REPORT RESUMES

ED 014 170 24 EC 000 522

DEVELOPMENT OF TWO STANDARDIZED MEASURES OF HEARING FOR SPEECH BY CHILDREN.

BY- SIEGENTHALER, BRUCE M. HASPIEL, GEORGE S.
PENNSYLVANIA STATE UNIV., UNIVERSITY PARK

REPORT NUMBER CRP-2372 PUB DATE 30 JUN 66
REPORT NUMBER BR-5-1064
CONTRACT OEC-5-10-003
EDRS PRICE MF-$0.50 HC-$5.16 127P.

DESCRIPTORS- *AUDITORY TESTS, *AUDITORY DISCRIMINATION, *STANDARDIZED TESTS, *TEST CONSTRUCTION, CHILDREN, TEST RELIABILITY, THRESHOLD BY IDENTIFICATION OF PICTURES TEST, DISCRIMINATION BY IDENTIFICATION OF PICTURES TEST

USING PREVIOUSLY DEVELOPED FORMS OF THE DISCRIMINATION BY IDENTIFICATION OF PICTURES (DIP) TEST AND THE THRESHOLD BY IDENTIFICATION OF PICTURES (TIP) TEST, THE INVESTIGATORS SOUGHT TO IMPROVE THE MATERIAL AND PROCEDURE AND TO STANDARDIZE THE TESTS ON NORMAL HEARING CHILDREN. BOTH TESTS REQUIRED CHILDREN TO IDENTIFY PICTURES UPON HEARING NAMES OF PICTURES. THE TIP TEST (TWO FORMS) AND THE DIP TEST (THREE FORMS) WERE ADMINISTERED TO 295 CHILDREN AGED THREE TO EIGHT YEARS. RETESTS FOR RELIABILITY WERE MADE ONE WEEK LATER. CONCLUSIONS FOR THE TIP TEST WERE--(1) IT PRODUCES NO DIFFERENCES BETWEEN FORMS A AND B IN VARIABILITY OF THRESHOLD, IN THRESHOLDS BETWEEN SEXES, OR THRESHOLD VARIABILITY BETWEEN SEXES OR AMONG AGE GROUPS TESTED, (2) FORM B PRODUCES MORE INTENSE THRESHOLD MEASURES THAN FORM A, FOR AGES TESTED AND BOTH SEXES, (3) AN AGE EFFECT EXISTS FOR BOTH MALES AND FEMALES AND FOR FORMS A AND B, (4) TEST THRESHOLD RELIABILITY BETWEEN TEST AND RETEST IS ABOUT THREE DECIBELS, (5) INTELLIGIBILITY CURVE IS ESSENTIALLY EQUIVALENT FOR FORMS A AND B, FOR BOTH SEXES, AND FOR ALL AGE GROUPS TESTED. CONCLUSIONS FOR THE DIP TEST WERE--(1) THERE ARE NO SCORE DIFFERENCES FOR ANY FORM AT ANY OF THE PRESENTATION LEVELS USED, NOR IN VARIABILITY OF SCORES AMONG TEST FORMS FOR AGE OR SEX GROUPS, NOR IN LEVEL OF TEST SCORE AS A FUNCTION OF SEX, (2) SCORES INCREASED WITH AGE OF SUBJECTS, (3) STANDARD DEVIATION OF DISTRIBUTION OF ABSOLUTE DIFFERENCE BETWEEN TEST AND RETEST SCORES ON ALL FORMS IS APPROXIMATELY FIVE ITEMS, AND (4) TEST SCORE PLOTTED AGAINST PRESENTATION LEVEL OF ALL FORMS, BOTH SEXES, AND ALL AGE LEVELS TESTED IS 1.8 ITEMS CHANGE IN SCORE PER DB CHANGE IN PRESENTATION LEVEL FOR RANGE SRT TO DB THROUGH SRT PLUS 10 DB. APPENDIX CONTAINS DESCRIPTIONS OF BOTH TESTS AND INFORMATION ON ADMINISTRATION, SCORING, AND INTERPRETATION OF SCORES. REFERENCE LIST CITES 59 ITEMS. (JB)

ED014170

EC 000 527

FINAL REPORT

Project No. 2372

Contract No. OE-5-10-003

DEVELOPMENT OF TWO STANDARDIZED MEASURES OF HEARING FOR SPEECH BY CHILDREN

30 June 1966

U. S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

DEVELOPMENT OF TWO STANDARDIZED MEASURES OF HEARING FOR SPEECH BY CHILDREN

Project No. 2372
Contract ~~Grant~~ No. OE-5-10-003

Bruce M. Siegenthaler
and
George S. Haspiel

30 June 1966

The research reported herein was performed pursuant to a contract with the Office of Education, U. S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

The Pennsylvania State University

University Park, Pennsylvania

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ACKNOWLEDGEMENTS

The authors wish to recognize the many individuals who contributed to this research.

Specifically, we wish to thank the children and parents of the State College, Pennsylvania, area who so willingly volunteered to participate in the study, who made the many necessary personal arrangements required for participation in the project, and who endured the inconveniences associated with that participation.

We also wish to recognize the contributions of the administrators of the State College Area Schools who made available school records to us as sources of children's names.

The authors are grateful to the various nursery schools in the area who contributed young children to the study. These include the Pennsylvania State University Nursery School Program, the Cooperative Nursery School, the Jack and Jill Nursery School and the Gladys Carter School, all of State College, Pennsylvania. In addition, the preschool programs of the Grace Lutheran Church and Hillel Foundations of State College made subjects available to the study. Some children came from among the graduate student

families living in graduate housing facilities on the Penn State campus.

The artists for the pictures, especially Joe Cervello, and the director of the Graphic Services at Penn State, Paul Stevenson, were patient, tolerant, and imaginative in the generation of the test pictures. Our thanks to them for their understanding help.

The several graduate students at Penn State who served as research assistants worked hard and long on carrying out many of the hearing test procedures, and they are most deserving of our thanks.

A special note of appreciation is due Dr. Jack Sparks, Assistant Professor in the Department of Educational Psychology at the Pennsylvania State University. Dr. Sparks gave most generously of his time and skills as our statistical consultant. His advice and help in statistical computation resulted in imaginative and meaningful analyses which enriched the study.

B.M.S.
G.S.H.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS ii

LIST OF TABLES v

LIST OF FIGURES vii

CHAPTER

I. INTRODUCTION 1
II. REVIEW OF THE LITERATURE 3

Threshold Testing 3
Discrimination Testing 8

III. PROCEDURES 18

Subjects 18
Development of Test Pictures 23
Experimental Test Environment 30
Test Series 35
Data Tabulation 37

IV. RESULTS AND ANALYSIS 38

Analysis Related to TIP Test 38

Comparisons between TIP Forms A and B 38
Comparisons between the Sexes 42
Comparisons among Age Groups......... 45
TIP Threshold Measurement Reliability .. 51
TIP Intelligibility Curve 54

Analysis Related to DIP Test 56

Comparisons among DIP Forms 56
Comparisons between the Sexes 65
Comparisons among Age Groups 69
DIP Test Reliability 76
DIP Score Discrimination Curve 81

V. CONCLUSIONS AND DISCUSSION 83

Conclusions Regarding TIP Test 83
Conclusions Regarding DIP Test 85
Comments Regarding Experimental Findings 86

VI. SUMMARY 91

REFERENCES 95

APPENDIX A: Protocol for TIP and DIP Tests 100

LIST OF TABLES

Table Page

1. Numbers of children by age categories available and tested 22

2. Word-pictures on the TIP test cards 24

3. Word-pictures appearing on DIP test cards 27

4. Call test word protocol for each form of DIP test 28

5. Transient sound levels for control and test rooms, and ASA standard for audiometric test rooms 31

6. Mean TIP test scores for test and for retest by age and sex 39

7. Summary of Lindquist type VI analyses of variance for TIP test data 40

8. Comparisons between TIP Form A and Form B mean thresholds by age and sex groups 41

9. Comparison between males and females for mean TIP test threshold scores within yearly age groups 44

10. C test values for differences among yearly age groups in mean TIP initial test thresholds, males only 46

11. C test values for differences among yearly age groups in mean TIP initial test thresholds, females only 47

12. Product-moment correlations between TIP initial test and retest thresholds 52

13. Range of differences, and absolute mean difference in decibels between test and retest for TIP thresholds . 53

14. Mean DIP scores for age and sex subgroups on the initial test at each presentation level 57

15. Mean DIP scores for age and sex subgroups on the retest at each presentation level 58

16. Summary of Lindquist Type III analysis of variance for DIP at level SRT + OdB, Forms 1, 2, 3 59

17. Summary of Lindquist Type III analysis of variance for DIP at SRT + 5 dB, Forms 1, 2 and 3 60

18. Summary of Lindquist Type III analysis of variance for DIP at level SRT + 10 dB, Forms 1, 2 and 3 61

19. Standard deviations of distribution of DIP scores for age and sex groups at each presentation level, initial test . 63

20. Standard deviations of distribution of DIP scores for age and sex groups at each presentation level, retest 64

21. Mean DIP test scores combined areas Forms 1, 2 and 3 for males and for females at each age group, and at each presentation level 66

22. Comparison between males and females within yearly age groups for DIP Test at levels SRT + 0dB, and + 10dB, using Dunn's multiple comparison procedure ... 67

23. Dunn's C test values for differences among yearly age groups in mean DIP test scores at level SRT + 0dB, males only 70

24. Dunn's C test values for differences among yearly age groups in mean DIP test scores, at level SRT + 5dB, males only 71

25. Dunn's C test values for differences among yearly age groups in mean DIP test scores, at level SRT + 10dB, males only 72

26. Dunn's C test values for differences among yearly age groups in mean DIP test scores, at level SRT + 0dB, females only 73

27. Dunn's C test values for differences among yearly age groups in mean DIP test scores, at level SRT + 5dB, females only 74

28. Dunn's C test values for differences among yearly age groups in mean DIP test scores at level SRT + 10dB, females only 75

29. Product moment correlations at each presentation level between initial test and retest scores on DIP for males and females combined (N 295) 79

30. Range of differences and absolute mean differences between test and retest for DIP scores 80

LIST OF FIGURES

Figure Page

1. TIP test protocol and scoring sheet 25
2. Mean TIP test scores by age group - males 49
3. Mean TIP test scores by age group - females 50
4. Mean DIP test scores by age group - males 77
5. Mean DIP test scores by age group - females 78

CHAPTER I

INTRODUCTION

Good educational practice requires that each child be given adequate opportunity to develop to his fullest potential. The widespread practices of administering hearing tests periodically to school children, follow-up otological and audiological procedures, special educational programs for the hard-of-hearing child, and schools for the deaf all testify to the generally recognized need for giving special attention to the child with a hearing loss. The clinical audiologist often has the responsibility for detecting educational and other problems related to hearing disorders, for his area of professional interest is at the core of a wide range of communication, social, and educational problems in hypoacousic children. To discharge his responsibility, the audiologist needs adequate clinical tools for the detection of hearing loss in children.

The hearing evaluation of children in the age range three to eight years of age presents special problems. Although pure tone audiometric techniques have been well developed for children, good hearing tests for speech are not available. The standard W-1, W-2, and W-22 tests clearly are not suitable for this age group, and clinical experience has demonstrated shortcomings in the usual testing with phonetically balanced word lists such as PB-50, PBK and PBF. Although each audiologist usually devises some procedures for his own practice, there are no generally available standardized audiology tests for measuring speech-hearing in young children.

Included in the problems of speech-hearing testing for children are their fear of the clinical situation, their frequent reluctance or inability to give intelligible oral response to test items, and the difficulty in finding test materials appropriate for their age level which have adequate acoustic and phonetic features.

The solution to these problems by the development of satisfactory audiological test materials for speech reception threshold and for speech discrimination would be a material contribution to the ability of the clinical audiologist to assess the hearing and associated educational problems of children, for often educational and other recommendations are dependent upon hearing status.

It was the purpose of the research reported herein to develop further and to standardize, on a group of normal hearing children, speech reception threshold and speech intelligibility test materials and procedures. The basic forms of both of the tests had been outlined by previous research (Siegenthaler, Pearson and Lezak, 1954; and Haspiel, 1961). Their potential, validity and reliability had been demonstrated by pilot studies. The present project included producing revised sets of test items and pictures, obtaining normative data as a function of age and sex for each test, estimating reliability of the two tests for normal hearing children, and preparing test procedure instructions for each test.

CHAPTER II

REVIEW OF THE LITERATURE

Although there are several facets of auditory function which describe an individual's ability to hear and understand speech, most workers in the field of audiology have accepted two dimensions of hearing, namely sensitivity and discrimination, as fundamental and requiring detailed investigation.

Threshold Testing

In the measurement of sensitivity we are concerned with determining thresholds, i.e., the least stimulus strength necesary for the subject to respond to a stimulus a given per cent of the time. Hirsh (1952) described two thresholds for speech measurement: the threshold of detectability, i.e., the intensity level at which the subject can detect the presence of speech 50% of the time; and the threshold of intelligibility, i.e., the stimulus level necessary to understand 50% of the speech presented. Usually the clinical task is to determine intelligibility rather than detectability for speech.

A number of standardized tests have been devised to measure the threshold of intelligibility. These tests use materials such as nonsense syllables, spondaic words, monosyllabic words, questions, or connected discourse. They generally meet the criteria that Carhart (1953) suggested for the measurement of any speech threshold, such as speech tests should contain items which are reasonably representative of connected speech and the items should be selected for homogeneity of audibility. Harris (1949) suggested a third criterion, that of homogeneity of item intelligibility to insure equal audibility of items.

Early tests of sensitivity included the whispered voice and spoken voice tests, which were crude quantitative tools for describing threshold in terms of distance from the speaker at which the listener could understand the test items. Because these tests did not use electronic amplification of speech, test items and procedure were difficult to standardize. Fowler (1941) recommended introducing a sound level meter into the test situation to permit the tester to monitor the intensity level of his speech. This helped to eliminate the inconsistent level of a given speaker's voice or inconsistent levels among various speakers. It also permitted the tester to remain in a constant position while varying his speech intensity levels rather than changing the distance of the listener from the source of the stimulus item.

While several tests exist for the measurement of speech sensitivity among adults, these are not appropriate for children. Adult tests do not take into account the special problems which arise in the testing of children. Ewing (1943) suggested that the stimulus employed must be meaningful to the child if he is to respond to it. Pure tones, for example, are seldom heard by the child in daily life and the child cannot be expected to respond to them in a meaningful way. He also stated that children who are placed in an unfamiliar environment such as a clinic soundproof room and required to perform an unfamiliar task might be fearful and hesitant.

Morley (1948) pointed out that the problem of short attention span is critical when testing the hearing of a young child. This is particularly important because of the length of the battery of tests commonly administered in performing a hearing aid evaluation or a complete audiometric examination for medical purposes. If any one test requires too much time, the results of other tests may be affected in an adverse fashion. The child is restless after a short testing period of time and may easily become fatigued.

A further problem which arises in some tests used with children is the difficulty of interpreting oral responses. Children with hearing impairment frequently exhibit articulation errors, vocabulary limitations and voice problems which make it difficult or impossible for the tester to understand their responses. The usual method for overcoming these procedural problems is to demand a simple motor (non-speech) response from the child.

Keaster (1947) devised an acuity test for speech using oral commands to eliminate some of these problems. Pictures were placed in various parts of the room and moved by the child upon the request of the examiner. It was assumed that the child understood the command if the response was appropriate. Failure to comply was evidence of the child's inability to hear the command. Although vocabulary was controlled for age level, the sentence length and the relatively complex language used sometimes proved difficult for the child with a communication handicap. Also, the assumption that a child did not hear or understand was sometimes in error because of lack of cooperation or fear which might also have caused the child to fail to follow directions.

Meyerson (1947) constructed a threshold test for children using thirty-six spondaic items together with pictures which represented them. The child was instructed to point to the picture corresponding to the stimulus word. In studying the responses of a group of fifty nursery school children, Meyerson found that there was a close relationship between a child's threshold for speech and his threshold of detectability as measured with pure tones. The conclusion was that a picture test using spondaic items was a simple reliable method for testing the hearing of young children. However, the test assumed that the spondee words were suitable for children as they were for adults, and the word-pictures were appropriate for younger children.

Monsees (1953) recorded a speech acuity test similar to Keaster's in vocabulary and the use of commands but employed objects rather than pictures. The same limitations apply to Monsees' procedure as to Keaster's. She also developed a "fading numbers" test (1953) in which pairs of digits were presented with progressive attenuation until threshold was established. Although the numbers used were speech items, they cannot be said to represent English generally. The numbers one through ten (without seven) which were used in this test contained a limited number of phonemes, which are easily guessed upon perception of the vowel component of the word.

Sortini and Flake (1953), also recognizing the difficulty of obtaining a verbal response or a voluntary action response to complex commands from children described a procedure used by many audiologists in their practice and requiring only the pointing to objects or toys to indicate hearing. This activity is a familiar one to children and does much to reduce the fear generated by the unfamiliar test situation. However, selection of objects (and their names) was not demonstrated to be based on factors other than interest to the child, e.g., test items did not have demonstrated suitable acoustic characteristics.

The Picture Identification Test (PIT) of speech hearing sensitivity reported by Siegenthaler, Pearson, and Lezak (1954) employed monosyllabic words represented by pictures. This test was based on earlier work by Pearson (1951). The words and pictures were validated with appropriate samples of normal and of hearing impaired children. The test thresholds have been shown to have satisfactory correlation with pure tone average threshold (Mullen, 1954, Siegenthaler and Strand, 1964). The only response required was pointing upon command. (The PIT is the prototype test which served as a basis for the present Threshold by Identification of Pictures Test).

Sims (1961) described a picture-identification test using spondaic words, which was employed to evaluate the hearing of preschool children. An analysis of the results indicated that the test was unsuitable for use with three-year old children and that considerable modification of the test was necessary if it were to be given successfully to four and five-year old children.

In summary, a number of children's hearing threshold tests have been devised, either by modifying existing adult procedures and materials or by attempting to use materials and procedures which satisfy the behavioral needs of children. Although such tests have been reported by their authors to be successful, none are in widespread use, and for some the test materials are not readily available.

The present authors and a number of their former students have been using successfully the procedure of having the child identify pictures by pointing, with the test procedures and materials based upon careful consideration of response modes of children, interest and attention span factors, and careful selection of test items for acoustic factors (Pearson, 1951; Siegenthaler, Pearson, and Lezak, 1954). A major portion of the work reported herein concerns further development of the test procedure for speech reception threshold measurement in children, together with deriving normal-child data for use as a base-line clinical standard.

Discrimination Testing

Discrimination in its broad sense can be described as the process whereby one selects a foreground signal from a group of other competing signals. In the measurement of speech discrimination ability we are concerned with the evaluation of the individual's ability to sort those verbal stimuli which are important from those which are unimportant, to identify them, and in some cases to relate them to other parts of a verbal statement. Although the detection of signals in noise is what seems to be suggested here, most clinical discrimination testing is done in the absence of noise (except where there is a specific interest in the S/N ratio). For the present research, discrimination testing, being done in quiet, can be considered as a special case of signal detection testing, with the listener's problem being that of recognizing the signal from among all those signals in that set which it might have been.

In this dimension of hearing function the tester wishes to describe the ability of the individual to identify speech units at a given intensity level rather than to measure the intensity necessary to understand 50% of the test items as is the case in threshold testing. Hirsh (1952) described the measurement of speech intelligibility as a supra-threshold process. That is, it is important to know not only how faint speech can be before it is no longer heard, but also how much speech can be understood when presented at a level overriding a hearing impairment. Such a statement of hearing function is essential if the individual is to be evaluated as a possible candidate for prosthesis or surgery designed to increase the loudness of sound reaching the inner ear. If the potential for speech discrimination is low this factor must be taken into account in evaluating the probable adequacy of any remediation procedure.

Modern speech discrimination testing originated in the work done at the Bell Telephone Laboratories by Fletcher and Steinberg (1929). Later work done at the Harvard Psycho-acoustics Laboratory furthered the knowledge and procedures used in discrimination measurement (Egan, 1944). Among the contributions of the Psycho-acoustics Laboratory was the presentation of phonetically balanced words as appropriate materials for measuring the ability to understand speech. The concept of distributing the phonemes of speech in the test materials (as in the PAL PB-50 word lists) to represent everyday speech is a reasonable one which satisfies the criteria of face validity for testing speech understanding. Although these materials were designed initially for evaluating communications systems, they were soon applied to audiological problems. Silverman and Walsh (1946), for example, described the use of the phonetically balanced word lists for audiological diagnosis with special reference to detecting the presence of inner ear involvement in otosclerotic ears as evidenced by a reduction in the maximum intelligibility score. The original PB word lists were revised by Hirsh et. al. (1952) by deleting items which were either too difficult or too easy to understand and by adjusting recording levels for fine differences in test words. New phonograph recordings were made of the PB word lists, and made available for discrimination testing. Carhart et. al. (1963) suggested further revision in PB word lists to permit them to be used with greater reliability.

Although there exist several tests for measuring auditory discrimination in children which are not dependent on a phonetic balance concept, these have largely been applied to children who have articulation problems rather than hearing impairments. Travis and Rasmus (1931), Hansen (1944), Mansur (1950), Templin (1952), Pronovost and Dumbleton (1953), and Shiefelbush and Lindsey (1958), among others, have developed such speech hearing measures. A number of these tests have used picture forms and have at times been employed with hearing impaired children, but not for the usual audiological purposes of discrimination testing.

Speech discrimination tests devised for use with children for the measurement of audiological impairments have been few in number. Modifications of the PB word lists were made by Hudgins (1949) and by Haskins (1949) for use with children who might have difficulty in handling the items in the PB-50 word lists. Sims (1961) developed a multiple choice picture test with phonetically balanced monosyllabic words. This test was not found to be useful in its existing form and further modifications were suggested before it was considered appropriate for preschool children. Nasca (1964) also described a picture test of speech discrimination which was not based on phonetic balance and found that it had merit for application with young children with sensori-neural losses and concommitant language delay.

Most of the above described tests were based on the concept of phonetic balance originally suggested as the important dimension in the construction of a discrimination test. However, there are indications that phonetic balance is not a critical factor in the measurement of speech discrimination. Half word lists drawn either randomly or systematically from PB-50 lists give essentially the same intelligibility scores as do full fifty-word lists (Reznick, 1962; Tobias, 1965). Goldstein (1965) systematically eliminated the most intelligible and least intelligible words in PBK word lists and found that nearly constant intelligibility scores were obtained although the lists were reduced to as few as fifteen words. This was despite the fact that the original phonetic balance of the PBK lists was no longer maintained.

Even though there is accumulating evidence that phonetic balance is not vital or necessary in measuring speech discrimination, to date few useful tests for speech discrinimation based on other factors have been devised.

There are suggestions from the literature which point to acoustic parameters which are promising for the construction of speech discrimination tests. Karlin (1942) using a factor analysis technique isolated several factors which appeared important, including pitch, quality, auditory-perception, auditory resistance to distortion, and auditory span formation. Mange (1953) showed that the ability to synthesize speech elements may be critical in the understanding of speech. Hanley (1956), in a factor analysis of a battery of twenty-six hearing tests, indicated eight such variables which might be related to the understanding of speech. Solomon et.al. (1960) too found several similar factors which were critical in speech understanding capacities, including auditory discrimination.

Black (1952) described familiarity of items as a major determinant of word intelligibility, as did Howes (1957), Rosenzweig and Portman (1957), Oyer and Doudna (1960), and Owens (1961).

Exposure to speech materials prior to testing has been demonstrated to be a significant factor in increasing intelligibility (Egan, 1944; Mason, 1945; Thwing, 1956). Furthermore, a message is more easily understood when selected from a small group of messages known in advance than when selected from a large message set, indicating too that the probability of occurrance directly affects the possibility of correct interpretation of a statement. (Miller, Heise and Lichter, 1951; Spilka, 1954).

A particularly important set of acoustic parameters of speech has been identified as being basic to speech intelligibility and discrimination. Potter, Kopp and Green (1947) observed that sound spectrograms could be interpreted visually by giving special attention to the factors which they called voicing, pressure pattern and influence. That is, spectrograms of syllables, words, and sentences could be recognized by observing (a) presence or absence of voiced phonemes, (b) continuant-plosive-affricate-glide acoustic patterns of consonants, and (c) transitions between adjacent phonemes. Joos (1948) also reported that the same factors were critical to the recognition of speech, using the sound spectrograph as the analyzing instrument.

Voicing may be described in acoustic terms as the presence or absence of phonation. A large number of pairs of consonants in our phoneme system vary principally in this factor: p/b, t/d, k/g, f/v, s/z, th (voiced)/th (unvoiced), and ch (unvoiced)/ch (voiced). This is a convenient variable to measure because other phonetic factors are relatively constant between the phoneme pair.

A second phonetic factor, influence, was described by Potter, Kopp, and Green (1947) as the assimilation effect between adjacent phonemes, with the effect especially observable between consonant and vowel, and with the effect of consonant especially strong upon the vowel rather than vowel upon consonant. It may be observed by measuring the frequency change of the formant bands representing the vowel, but it is especially evident in the second formant. These authors state the following:

> These influences appearing in the patterns, are produced when the articulators change from positions characteristic of one sound to positions characteristic of the succeeding sound. ...The basis of how sounds influence each other is in the way resonating cavities change shape and size as we say one sound after another in pronouncing words. (1947, p. 38)

In describing the importance of influence to speech perception Joos stated:

> ...the most that a listener can get out of a consonantal segment itself is just barely enough to classify it as voiced or voiceless, stop, fricative or resonant. The rest of the evidence which classifies the consonant as to PLACE of articulation, the listener finds only within the confines of contiguous vowels (1948, p. 123)

The third factor, that of pressure pattern, may be described as the amount of acoustic energy present at given times during production of the phoneme. (In terms of spectrographic analysis) pressure pattern...refers to the presence or absence of energy during successive time units during the time span of the speech sound. (Potter, Kopp and Green, 1947, p. 51)

Siegenthaler (1954) reported on the ability of listeners to perceive these phonetics factors. In his procedure, words were paired to be different in each of the phonetic factors and in combinations of them and subjects were required to distinguish them. His finding was that there were wide differences between individuals, not related to aural pathology, in the perception of phonetic factors. A factor analysis led to the conclusion that there was a general speech perception factor related to perception of these phonetic elements. It was also concluded that the test using the phonetic variables did not indicate the same dimensions of hearing as did pure tone audiograms, speech thresholds, or otological diagnosis.

Over the past fifteen years researchers at the Haskins Laboratory have investigated a number of elements contributing to the perception of speech. Their studies verify the importance of transitional factors, manner of articulation (i.e., whether the sound is a plosive or a continuant) and voicing as critical components (Cooper, et.al., 1952; DeLattre, et. al., 1955; Liberman, et.al., 1956, 1959).

Recently Peters (1963) obtained data supporting the importance of these factors. His work did not utilize the sound spectrograph, but centered on perceptual judgments. Listeners were required to hear pairs of speech sounds and to indicate closeness or distance between them on a subjective scale. Utilizing the concept that auditory space has several dimensions leading to intelligibility, Peters concluded that manner, voicing, and place of articulation are of importance in this respective order (i.e., phonemes are first sorted according to manner, voicing is next in importance, and place for some individuals is also important). In Peters' work manner of articulation referred to the plosives-continuant differential, voicing referred to presence or absence of laryngeal tone, and place referred to the place of articulated constriction in the oral region for the production of specific consonants. (The factor of place determines the transition which occurs between adjacent phonemes; transitions are displayed prominently as second formant changes in sound spectograms.)

Håspiel's Picture Identification Test of Discrimination for Speech (1961) used the variables of voicing, pressure pattern and influence in a test suitable for evaluating the speech discrimination of young children. His data indicated that the procedure has satisfactory reliability, and it discriminated among individual listeners in the age range eight to twelve years who had a variety of audiological and otological conditions. This test served as the prototype for the present test, Discrimination by Identification of Pictures (DIP).

In summary, audiological speech discrimination tests presently in use have been formulated with relatively little concern for the particular problems of hearing impaired children. Furthermore, most tests have been based on phonetic balance which in light of recent research might more properly be replaced by several experimentally suggested acoustic-phonetic factors.

In the research reported in the following, a test procedure adapted for young children according to the features mentioned earlier, and incorporating phonetic factors believed to be important to speech intelligibility is described and evaluated. The research effort was directed toward further developing the test materials and protocol and studying the responses of normal children, to arrive at clinically useful normative data.

CHAPTER III

PROCEDURES

Subjects

The design of this study called for approximately three hundred normal children between the ages of two years-ten months and eight years-three months inclusive, divided at approximately equal age intervals, with approximately equal division between the sexes at each age group.

The following steps were taken to obtain the subject sample, utilizing the population of the State College, Pennsylvania area:

1. Contact was established with school administrators of the College Area Schools of State College, Pennsylvania.
2. The school records of all children registered in the College Area School beginning in the Fall of 1964, through the age of eight years, were made available to the project staff. Total school enrollment, grades K through 12 was approximately 6700. Dr. William E. Babcock, Director of Instruction for the State College Area School District was especially helpful in arranging access to the school records.
3. The school record for each child was examined, and those children whose records indicated the following were selected as being eligible for the project·
    a. Exact birthdate.
    b. Evidence that the routine school hearing test had been passed successfully on each occasion; absence of speech problem.
    c. School psychological tests (usually group tests such as Otis) indicated IQ between 90 and 110.
    d. Had not repeated a school grade.
    e. Free from known visual, neurological or educational problems.
4. The children pre-selected by steps a. through e. above were further screened by their teachers. That is, the

teacher of each child was asked to verify or deny the evaluation of "normal" for each child. If in the teacher's opinion a child was clearly superior or inferior in school achievement or adjustment, he was eliminated.

5. A projected testing schedule was arranged and eligible children were categorized by monthly age so that at the time of testing, according to the projected schedule, they would be within three months of the yearly age groups.

6. Local newspaper publicity was obtained to prepare the community for the contacts to follow.

7. Of all children eligible for testing, thirty boys and thirty girls for each of six age groups were randomly selected for initial contacts, limiting the selection to children residing within the State College, Pa. borough limits.

8. A letter was written to parents of each child selected requesting participation in the project, and asking for a reply.

9. To find children in the pre-school age categories the Director of the Pennsylvania State University Nursery Schools was contacted, as were directors of the Cooperative Playschool, Jack and Jill School and the Gladys Carter School, all of State College.

10. Because of the relatively smaller number of young children available for the study through these channels, the following additional contacts were made to obtain a larger pool of pre-school age children:

    a. Department of Housing, The Pennsylvania State University, with the request that names of families having pre-school children living in graduate student housing facilities of the University be made available. Contacts were established with these families. Mr. Otto Mueller of the University Department of Housing was very cooperative in this effort.

    b. Bellefonte, Pa. (the nearest community to State College) Area Schools kindergartens.

    c. Nursery school programs operated as part of the Sunday Schools of Grace Lutheran Church and the Hillel Foundation, both of State College, Pa.

11. The teachers in the nursery schools were asked to preselect children believed to be of normal-average intelligence and social-emotional adjustment (neither inferior

nor superior) according to their observation and professional judgement. Children were selected randomly for further contacts from among the names provided. The families of children were reached in a manner similar to the public school children. The subjective teacher rating of "normal" as for pre-school children was also used for nearly all kindergarten children obtained both in State College and Bellefonte because school intelligence tests were not yet done on these children. The project staff interviewed parents and observed child behavior for the subjects drawn from the University housing facility.

12. Upon the receipt of written indication from the parents that they would cooperate with their child in the study, they were notified that specific appointments would be made at a later date. As a child approached the age level of the group in which he would be included, telephone calls were made to the parents to arrange for final case selection and experimental testing.

13. In those age categories where an inadequate number of parents volunteered their children, the residual pool of eligible children was tapped for replacements, using random selection, until the final testing groups were obtained. (In some cases children became ill, were found to have developed hearing loss or ear pathology, or the family left the area after volunteering for testing. The attrition rate was especially high among the younger children. Where possible these lost subjects were replaced from the pool of potential subjects, using random selection.) For this replacement of subjects, eligible children residing outside the State College, Pa. Borough limits were used, but all were within ten miles travel distance, and all except the pre-school subjects attended the College Area Schools.

14. A subject to be included in one of the yearly age interval subgroups was scheduled for testing so that at the time of testing he was plus or minus three months of the year. That is, a three-year-old group child had to be between the ages of 2-10 and 3-3 inclusive; a four-year-old group child had to be between the ages of 3-10 and 4-3 inclusive; and so forth. The parents were instructed that at the time of the child's appointment he was to go to the University Health Service for otological examination. Taxi transportation was provided for mother and child (as well as other children in the family if they had to accompany the mother). At the end of testing, taxi fare was provided to return home for the family. A similar arrangement was made for the return visit for retesting.

15. The otological examination was done by a staff physician at the University Health Service. The examination of both ears included inspection of the external auditory meatus and the eardrum by otoscope, oral examination with specific reference to enlarged or diseased tonsils and adenoids, and nasal inspection with questioning of parent with reference to allergic conditions. The examination results were reported on a check-sheet and carried by the parent to the audiological test immediately following.

16. At the time of the initial audiological testing, a case history interview was carried on with the parent. The purpose was to assure that the parent did not consider that the child had a hearing loss or any hearing difficulty, that the child was not believed to have a current medical condition of the ear (even though the immediately preceding otological examination was negative), that the child was free of a speech problem with normal speech development for his age, to affirm that there appeared to be no developmental lags, and that for school age children a school grade was not repeated.

17. If the preceding procedures (otological examination and case history) were negative, the child was given a pure tone threshold test in each ear, air conduction, for the octave frequencies 250 through 8000 cps in the sound room described later. The test was done using a descending, ascending, descending series of threshold estimates at 5 dB steps with the threshold taken as the lowest level at which at least two of three tone presentations were heard. Hand raising or play audiometry was used, as appropriate for the child. A Maico F-1 audiometer was used, and given weekly calibration checks with a Bruel and Kjaer audiometer calibration system, model 158. No more than a five dB audiometer correction was needed at any time during the testing program at any frequency. The ASA-1951 standard calibration was used.

18. For a child to be accepted for experimental testing, all thresholds for both ears had to be 10 dB or better.

19. According to the above procedures the final sample of subjects was obtained. Table 1 gives data regarding numbers of subjects involved in these subject selection procedures, and final sample size.

Table 1. Numbers of children by age categories available and tested.

Age Category	Eligible For Testing	Initially Contacted	Obtained As Replacements	Total Tested	Mean Age In Months
3 Year Group					
Male	35	30	5	17	35.8
Female	41	30	7	15	37.0
4 Year Group					
Male	67	30	8	20	48.2
Female	72	30	6	24	47.4
5 Year Group					
Male	81	30	4	24	61.0
Female	74	30	2	16	60.7
6 Year Group					
Male	122	30	3	30	72.2
Female	143	30	4	30	72.5
7 Year Group					
Male	195	30	4	30	83.9
Female	212	30	5	30	84.5
8 Year Group					
Male	246	30	4	29	96.9
Female	342	30	3	30	96.1
Total Male	746	180	28	151	
Total Female	884	180	27	144	
Total	1630	360	55	295	

Development of Test Pictures

The tests utilized in this study were described in detail in previous literaure. However, the following gives a brief overview of each:

The Threshold by Identification of Pictures test (TIP) was originally described by Siegenthaler, Pearson and Lezak (1954) and was based upon earlier work by Pearson (1951). Form A of this test is composed of a set of six cards each carrying five pictures per card in color (for example, fish, dog, house, ball, comb). The specific test items were chosen by previous research for familiarity to children, to be unambiguous in name and to have specific degrees of audibility. All appear among the first five hundred words of A Basic Vocabulary for Elementary School Children (H.A. Rinsland, New York, MacMillan, 1945). The first card is for practice, and the remainder are test cards. The child is presented one card at a time, and told "Point to the ..." according to the prepared test protocol using the speech audiometer. Five responses are obtained using the pictures on the card. After each item the tester's voice is reduced 5 dB. The first item on each card is at the original beginning level, 10 dB above estimated threshold. The test obtains 25 responses from the five cards, and permits obtaining of dB level for a 50% threshold using either the graphic or the tabular method. Form B protocol is the same as Form A, but the picture items are different. The call words for each card were selected randomly from among the five words available for the card. In all cards but one in the entire set of cards for Form A and Form B, four different words per card were called, with one word repeated in each card. Table 2 lists the different word-pictures on each card of the TIP Test. Figure 1 is the test protocol for the TIP Test.

The Discrimination by Identification of Pictures test (DIP) was described by Haspiel (1961). The DIP Test is composed of four practice and 48 test cards with two pictures per card. The subject is told to indicate one of the items on each card by being asked "Point to the ..." according to a prepared test protocol, using the speech audiometer. After each item the card is turned exposing the next pair of pictures. All words are called at the same intensity level, as predetermined by test conditions. The pictures are of things familiar to children, unambiguous in name, and arranged in pairs to be different in the phonetic factors of consonant voicing, (e.g., pear-bear), of transition, (e.g., peas-keys), of pressure pattern of consonants, (e.g., hat-cat), or combinations of these phonetic factors. All words appear in The Teacher's Word Book of Thirty Thousand Words (Thorndike and Lorge,

TABLE 2. Word-pictures on the TIP test cards.

TIP Form A

Practice Card A	Card A-1	Card A-2	Card A-3	Card A-4	Card A-5
bread	dog	house	top	mouse	key
cap	hand	bus	horse	comb	train
chair	cake	dress	ball	cup	milk
car	stove	kite	cat	clock	eyes
shoe	truck	flag	fish	drum	spoon

TIP Form B

Practice Card B	Card B-1	Card B-2	Card B-3	Card B-4	Card B-5
shoe	plane	gum	fish	clown	blocks
car	cow	hand	lamb	skates	watch
swing	pie	hat	top	man	doll
chair	socks	boat	frog	comb	tree
dog	door	bed	corn	gun	knife

THRESHOLD BY IDENTIFICATION OF PICTURES (TIP) SCORING SHEET

NAME ____________________ SEX ______ DATE __________ EXAMINER __________

ADDRESS ______________________________ BIRTHDATE ______________

Field _____ Earphones _____ Test Validity _____ Remarks ______________________

FORM A

Practice A	audiometer setting	A-1	A-2	A-3	A-4	A-5	NUMBER CORRECT
chair	____________	hands	house	horse	clock	eyes	______
shoe	____________	dog	flag	ball	mouse	train	______
chair	____________	truck	dress	cat	drum	milk	______
bread	____________	stove	bus	fish	cup	eyes	______
cap	____________	hands	kite	ball	clock	spoon	______
					TOTAL CORRECT		______

THRESHOLD

Right_____dB

Left _____dB

Bin. _____dB

FORM B

Practice B	audiometer setting	B-1	B-2	B-3	B-4	B-5	NUMBER CORRECT
shoe	____________	door	bed	lamb	skates	tree	______
chair	____________	pie	boat	corn	clown	knife	______
car	____________	cow	hand	top	comb	doll	______
swing	____________	socks	gum	lamb	gun	tree	______
dog	____________	pie	bed	frog	clown	blocks	______
					TOTAL CORRECT		______

SCORING GRID

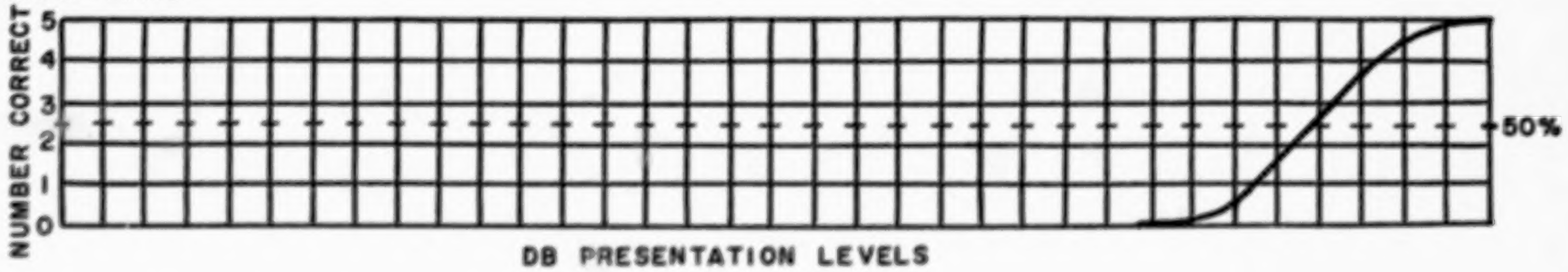


Figure 1. TIP test protocol and scoring sheet.

Teacher's College, Columbia University, New York, 1944) among the "most familiar" category. For each item the call word was selected by chance for DIP List 1. DIP List 2 is composed of the other call word for each pair. DIP List 3 was determined by randomly selecting another time a call word for each card. Scoring is according to correct or incorrect selection of the item called in each pair, and provides an overall per cent correct for each form of the test. Only the forty-eight test items-pictures are scored. Table 3 gives the words-pictures for each card. Table 4 is the call word test protocol for each form of the DIP Test.

Original sketches and sample pictures had been prepared previously for each of the tests. The Graphic Services Department of The Pennsylvania State University produced new and original drawings of each picture whose name appeared in the tests. (Paul Stephenson served as supervising artist for the project.) Preliminary color drawings were prepared for each different word, and these were tinted to approximate final coloration. After staff evaluation and improvement of the pictures, they were taken to the University Nursery School for screening with children.

Twenty-five three and four-year-old children in the nursery school viewed the pictures individually. Each picture was shown to the child and he was asked to name it. If a given picture was given the same name (the word intended to be illustrated) by all twenty-five children, it was accepted.

Any picture not correctly identified according to the test call word by any child was reconsidered. For a given child those pictures to be reconsidered were put two at a time in a group together with three pictures previously identified by the child. As the child viewed this group of five pictures the researcher asked

TABLE 3. Word-pictures appearing on DIP test cards.

Card	Word Pair		Card	Word Pair	
Practice A	cat	dog	VI-23	pot	dot
Practice B	chair	boat	VI-24	cone	bone
Practice C	key	kite	VI-25	key	bee
Practice D	kite	coat	VP-26	nail	sail
			VP-27	pen	men
V1	bear	pear	VP-28	gun	sun
V2	dear	tear	VP-29	feet	beet
V3	peas	bees	IP-30	bat	cat
V4	fan	man	IP-31	fire	tire
V5	goat	coat	IP-32	horn	corn
I-6	key	pea	IP-33	pear	hair
I-7	pup	cup	IP-34	log	dog
I-8	boat	goat	IP-35	fan	can
I-9	tea	pea	IP-36	cheese	peas
P-10	meat	beet	IP-37	shoe	two
P-11	saw	paw	VIP-38	light	kite
P-12	chain	cane	VIP-39	bees	cheese
P-13	seal	wheel	VIP-40	toes	rose
P-14	cheese	keys	VIP-41	rain	cane
P-15	ring	wing	VIP-42	bat	hat
P-16	bat	rat	VIP-43	thumb	gum
P-17	tail	sail	VIP-44	log	hog
VI-18	coat	boat	VIP-45	door	four
VI-19	toe	bow	VIP-46	can	man
VI-20	toy	boy	VIP-47	suit	boot
VI-21	tack	back	VIP-48	hair	bear
VI-22	cat	bat			

V-indicates voicing difference of initial consonants.
P-indicates pressure pattern differences of initial consonants.
I-indicates pressure pattern differences of initial consonants.

Cards 18-48 are of items differing in more than one phonetic factor, as indicated. All cards are numbered consecutively.

TABLE 4. Call test word protocol for each form of DIP Test.

Discrimination by Identification of Pictures (DIP)
Scoring Sheet

NAME ______________________________ AGE __________ DATE ______________

Card	DIP Form 1	DIP Form 2	DIP Form 3
Practice A	dog	cat	dog
B	chair	boat	chair
C	kite	kite	kite
D	coat	kite	coat
DIP V-1	bear	pear	pear
2	deer	tear	deer
3	bees	peas	peas
4	fan	man	man
5	coat	goat	coat
I-6	key	pea	key
7	pup	cup	cup
8	goat	boat	goat
9	pea	tea	tea
P-10	meat	beet	meat
11	saw	paw	saw
12	chain	cane	cane
13	wheel	seal	wheel
14	cheese	keys	keys
15	wing	ring	ring
16	rat	bat	rat
17	sail	tail	tail
VI-18	coat	boat	coat
19	bow	toe	toe
20	toy	boy	boy
21	back	tack	back
22	cat	bat	cat
23	pot	dot	dot
24	bone	cone	bone
25	bee	key	bee
VP-26	nail	sail	sail
27	men	pen	men
28	sun	gun	sun
29	feet	beet	feet
IP-30	cat	hat	hat
31	fire	tire	fire
32	horn	corn	horn
33	pear	hair	hair
34	dog	log	log
35	can	fan	can
36	peas	cheese	peas
37	shoe	two	shoe
VIP-38	light	kite	kite
39	cheese	bees	bees
40	rose	toes	rose
41	rain	cane	rain
42	hat	bat	bat
43	gum	thumb	gum
44	log	hog	log
45	four	door	four
46	man	can	can
47	suit	boot	suit
48	bear	hair	hair

the child "Point to the ...", randomly calling each of the five pictures, including those previously not named correctly.

The pictures incorrectly identified at least once among the twenty-five children during this reconsideration were redrawn and again submitted to all twenty-five children. This process of having pictures redrawn and re-evaluated was repeated until the imagination of the artists was exhausted. Some words were found to be especially difficult to illustrate. After approximately three to five drawings they were submitted to the twenty-five children again. Each child was taught the name of the picture by being given its intended name and being told about the qualities of the object or concept represented. After all of these difficult items were taught they were intermixed in groups of five, along with words already known by the child, and the procedure of "Point to the ..." for having the child identify the picture was followed. All of the difficult pictures were identified by the children after being taught them. In all, 28 words had to be redrawn at least once; six words had to be taught to at least one child.

The tinted drawings of the pictures were transfered into suitable form for the production of colored printing plates. Pictures were arranged in groups according to the requirements of each test card, the color plates were made and the pictures were printed by the offset process using an array of four colors plus black. In their final form, each picture card test measured 8-1/2 inches by 11 inches and each contained either five or two pictures depending upon the test for which it was made. Pictures were colored in subdued tones to minimize over-attractiveness.

The picture cards were arranged in convenient testing order and provided with covers and a plastic spiral binding. (The final book of test pictures together with protocols, Test Scoring Sheets, and guideline for test interpretation are contained in the sample copies submitted, and titled TIP and DIP Test Materials: A Manual of Procedure and Test Pictures, by Bruce M. Siegenthaler and George S. Haspiel.) Final binding is with spiral wire.

Experimental Test Environment

A room in the Chambers Building on The Pennsylvania State University campus was assigned to the project. This room was in a quiet part of the Chambers Building (built in 1961). It had approximately 190 sq. ft. of floor space. The room was equipped with a two-room IAC sound suite, series 600. The control room measured four by six feet and the adjoining subject sound room measured six by eight feet.

Acoustic measurements made within the sound rooms using a Bruel and Kjaer sound level meter and octave band filter attachment model 2203/1613 (audiometer calibration system set-up for free field measurements) indicated transient noise levels for octave bands within tolerance levels for audiometric test rooms according to the American Standards Association Standard S3.1-1960. Table 5 indicates the sound levels measured in the subject test room and control room at each band and the ASA Standard values, by octave bands.

The control room was outfitted with a Magnecorder PT6 tape

TABLE 5. Transient sound levels for control and test rooms, and ASA Standard for audiometric test rooms.

Octave Frequency Band	Levels in Test Room at Subject's Position. (Outer Room Blower Off. Test Room Blowers On.)	Levels in Control Room. (Outer Room Blowers Off. Control Room Blowers On.	ASA Standard S3.1-1960 Background for Audiometric Rooms.
62.5 cps	32.0	40.0	40
125 cps	25.0	29.0	40
250 cps	16.0	19.0	40
500 cps	15.0	18.0	40
1000 cps	14.5	17.5	47
2000 cps	15.5	17.5	57
4000 cps	16.5	18.3	67
8000 cps	18.0	16.5	67

Note: Levels in dB re .0002 dy/cm^2.

transport and amplifier, fed into a Grayson-Stadler Model 1160-A speech audiometer with a fifteen inch Jenson coaxial speaker. The speech audiometer was calibrated in sound pressure levels for field testing, reference .0002 dyne/cm^2 as measured on the C scale of a Bruel and Kjaer, Model 2203/1613 sound level meter. The audiometer calibration was monitored at weekly intervals and adjusted as necessary throughout the experimental testing. The test room was provided with an earphone to be worn by the tester. The tester also had a spring-return switch to control starting and stopping of the tape transport which fed the speech audiometer during experimental testing.

The subject room was equipped with a child size (adjustable) small table for holding the test materials, chairs suitable for the different age subjects and a chair for the tester. Experimental testing was done free field using tape recorded materials.

The outer room area provided waiting room for parents, desk work space, and record keeping and record processing areas.

Preparation of Experimental Test Sound Recordings

The call words for each picture card of the TIP test Form A and for Form B are as shown in Figure 1. For each column of five words there is a TIP card.

The call words for the TIP test were tape recorded using an Ampex 600 series tape transport and amplifier and an Electro-voice 633 dynamic microphone. Recording was done in a quiet room, with all words monitored at VU O. Each call word was preceded by the phrase "Point to the ...". The carrier phrase-stimulus word groups were read at six second intervals.

The call words for Forms 1, 2, and 3 of the DIP test were recorded in a like manner, with all words monitored to VU O.

The complete recording of all words appearing in the TIP and DIP tests was copied onto another tape using Magnecord PT6 equipment. For this dubbing note was made as to the volume unit meter peak level for the stimulus word, and by using an attenuator in the patch-cord circuit one to three dB adjustments were made in signal level to equalize more exactly the level of all stimulus words for a closer approximation to VU O.

This equalized tape was copied back on to the Ampex 600 series recording equipment to produce a master tape. The master tape was provided with a 1000 cycle calibrating tone, recorded at VU O for thirty seconds at the beginning of each of the TIP Forms and at the beginning of each of the DIP Forms. At the beginning of each TIP form the Form number was announced. After each carrier phrase-word item five decibels of attenuation was put into the circuit successively, so that the five words of a TIP group covered a twenty-five decibel range. At the end of a set of five words the attenuation was removed and the announcement, "Turn the page", was made. The next set of five words began at the original beginning level without attenuation and attenuation was again added after each word, in five decibel steps. TIP Form A and Form B were recorded in this manner. For DIP Forms 1, 2, and 3 the 1000 cycle calibrating tone at VU O was recorded at the beginning of each list of 52 carrier phrase-words. No successive attenuation was used during a DIP list recording. At the end of a DIP Form the announcement "End of DIP Form ____" was added. The master tape was recopied back onto the Magnecorder PT6 equipment several times to make a group of test tapes having desired sequencing of TIP and DIP forms, as required for subsequent testing. It was possible to record both TIP Forms and all three DIP Forms on a single 1200 foot reel of recording tape at 7-1/2 i.p.s.

Because the DIP tests were to be given at intensities relative to the SRT levels as determined by TIP tests, Form A and Form B were recorded at the beginning of a test tape, followed by DIP Forms. The order of use of TIP Forms A and B, and DIP Forms 1, 2, and 3 was rotated among subjects, but without DIP Forms 1 and 2 adjacent to each other. There were four test sequence orders as follows among the recordings:[1]

Sequence 1
- TIP A
- TIP B
- DIP 1
- DIP 3
- DIP 2

Sequence 2
- TIP B
- TIP A
- DIP 1
- DIP 3
- DIP 2

Sequence 3
- TIP A
- TIP B
- DIP 2
- DIP 3
- DIP 1

Sequence 4
- TIP B
- TIP A
- DIP 2
- DIP 3
- DIP 1

---

[1] DIP forms were administered at SRT + 0 dB, SRT + 5 dB, and SRT + 10 dB, with the dB levels rotated among subjects and sequence orders.

When a child was tested the 1000 cycle tone at the beginning of a given test list was monitored to VU O on the speech audiometer, and the starting level for a given test list was adjusted by use of the speech audiometer attenuators. Each test list for a child was checked for 1000 cycle tone calibration.

By use of the monitoring earphone by the tester who sat beside the child, and the over-ride on-off switch for the tape transport, it was possible to monitor stimulus words given to the child, to score the child's response, and to allow the tape to run continuously or at an interrupted pace according to the child's speech of response.

For SRT and discrimination testing the child was seated in the sound room at a measured distance from the loudspeaker to provide a calibrated speech signal. The test materials were placed before him on the table and the tester sat beside him with scoring sheet, earphone, and tape transport switch in hand. The child responded to a given test item by pointing to the picture of the word he thought he heard. During testing no item was repeated. A stand-by tape transport and smplifier unit were available if a tape unit did not operate properly.

Test Series

Each child was given the following experimental test series following subject selection procedures. The following test sequence is one of the four sequences used; a child received a TIP-DIP test sequence according to one of the orders mentioned earlier.

1. Both TIP Test Form A, and TIP Test Form B were administered. (SRT values were recorded in sound pressure level reference .0002 dyne/cm^2. For nearly all children the tabular method of scoring was possible. For the few for whom this was not possible, the graphic method was used for each test independently. The mean SRT across Form A and Form B was obtained. According to a predetermined and rotating test protocol, for half of the children Form B was administered first, while Form A was administered first for the other half.)

2. Short rest, out of the sound room.

3. DIP Form 1 at SRT plus 0 dB, and rest.

4. DIP Form 3 at SRT plus 5 dB, and rest.

5. DIP Form 2 at SRT plus 10 dB.

6. The child was scheduled to return within three days for retesting. In those few cases where a three day interval was not practical, the child was scheduled to be retested within one week of the initial test.

7. On the retest visit each child was given a pure tone air conduction screening test at 10 dB (American Standard-1951 calibration). The parent also was interviewed to report any observed change in the child's hearing since the initial test. (No child was accepted with reduced acuity for air conduction pure tones or with a parental report of hearing change between the first test and the return for retest. Some children became ill between test and retest, and had to be dropped from the study.)

8. The TIP Test Form A and Form B, and the DIP Forms 1, 2, and 3 were repeated for the child. The same sequence of test lists was administered for the retest as was used for the initial test. DIP Forms were administered at intensities relative to the mean of the TIP Forms A and B retest scores. (The order of DIP intensity level was rotated among subjects as on initial testing.)

9. Upon completion of testing, the parent was given a short verbal report. Later a report form was completed in a check list manner regarding the findings for each child, (including a pure tone air conduction threshold audiogram, otological findings, and experimental test scores) and sent to the parents.

Appendix A is taken from TIP and DIP Test Materials: A Manual of Procedures and Test Pictures, and gives detailed instructions for test administration and scoring.

Data Tabulation

Identifying data for each child, age to the nearest month at the time of testing, test list sequence, and TIP scores and DIP scores with indication of DIP presentation levels were recorded for each child on IBM tabulation sheets.

(In addition to the data of special interest to the present study, plans were made for data tabulation procedures which would handle a variety of audiological case types. The data tabulation form also specified the aural condition of the child as to medical diagnosis and the pure tone audiogram two frequency better ear average.)

These data for each child were punched on IBM cards and verified, using one card for identification data, one card for all TIP test and retest scores, and one card for DIP test and retest data.

This deck of cards, together with appropriate computer program cards, was entered into the Penn State Computer Center procedures for data analysis.

CHAPTER IV

RESULTS AND ANALYSIS

Analysis Related to TIP Test

Table 6 shows the mean TIP test scores obtained. The *Test* columns are the scores obtained on initial session with the child, and the *Retest* columns are the scores obtained on the second session with each child. All scores are in decibels reference sound pressure level .0002 dyne cm^2. The mean standard deviations across all subjects, shown in the bottom row of the Table, were obtained for each column as the arithmetic mean of the standard deviation values in the column.

As the initial analysis of the TIP test data a Lindquist type VI analysis of variance (Lindquist, 1953) was done. Table 7 is the summary of that analysis.

Comparisons between TIP Forms A and B. In Table 7 the significant Test-retest x Forms interaction indicates that the difference between TIP Forms differ among age groups. Further, it is seen in Table 7 that the test Forms F ratio was significant at the .01 level, and it may be concluded that there was a difference between test forms among the means indicated in Table 6. To test for the source of the significant difference in TIP test Form means, repeated computations of Dunn's multiple comparison procedure (Sparks, 1963) were done. (The Dunn multiple correlation procedure, resulting in C, is an appropriate and valid substitute for the t test.) Table 8 shows the mean differences between Form A and Form B threshold

Table 6. Mean TIP test scores for test and for retest by age and sex (dB SPL re .0002 dyne/cm^2).

Subject Group		Test				Retest			
		Form A		Form B		Form A		Form B	
		Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
3 Year Total		17.52		17.15		17.96		18.82	
Male	(N 17)	17.28	5.25	18.06	5.09	16.50	5.01	17.39	3.92
Female	(N 15)	17.75	5.26	17.17	6.05	19.42	7.13	20.25	5.99
4 Year Total		17.26		19.15		17.39		18.64	
Male	(N 20)	17.05	6.23	17.71	6.20	17.00	6.28	17.67	5.01
Female	(N 24)	17.50	5.66	20.58	6.16	17.78	3.68	19.62	4.67
5 Year Total		13.44		15.24		13.56		14.95	
Male	(N 24)	13.46	4.03	14.72	4.23	14.09	4.27	15.00	5.55
Female	(N 16)	13.43	6.02	15.77	5.70	13.03	5.66	14.90	5.67
6 Year Total		13.48		15.10		13.40		15.36	
Male	(N 30)	12.29	3.59	12.71	3.69	11.67	4.03	13.40	4.30
Female	(N 30)	14.68	4.08	17.49	4.84	15.13	4.52	17.33	4.78
7 Year Total		11.20		13.08		10.78		12.28	
Male	(N 30)	10.45	5.40	12.32	4.33	10.57	5.40	12.55	4.58
Female	(N 30)	11.95	4.44	13.84	5.10	11.00	4.64	12.00	4.56
8 Year Total		9.00		10.55		8.02		10.08	
Male	(N 29)	8.62	6.94	9.83	7.98	7.23	5.09	9.65	4.85
Female	(N 30)	9.37	4.81	11.27	5.32	8.81	5.41	10.52	5.50
Mean Std. Dev. Across All Subjects			5.14		5.39		5.09		4.95

Table 7. Summary of Lindquist type VI analysis of variance for TIP test data.

Source	Sums of Sqs.	d.f.	Mean Sq.	Error Term for F	F	P
Between Subjects	35198.280	292	120.542			
Age-Sex	12398.290	11	1127.117	B	13.891	.01
Error-B	22799.990	281	81.139			
Within-Subjects	8123.760	879	9.242			
Test-Retest	18.880	1	18.880	1 (w)	1.256	N.S.
Test-Retest by Age-Sex	171.680	11	15.607	1 (w)	1.039	N.S.
Test Forms	707.810	1	707.810	2 (w)	134.968	.01
Forms x Age-Sex	118.590	11	10.781	2 (w)	2.056	.05
Test-Retest x Forms	0.020	1	0.020	3 (w)	0.041	N.S.
Test-Retest x Forms x Age-Sex	47.650	11	4.332	3 (w)	0.893	N.S.
Error (w) Total	7058.950	843	8.374			
Error 1 (w)	4222.750	281	15.028			
Error 2 (w)	1473.640	281	5.244			
Error 3 (w)	1362.560	281	4.849			
Total	43322.040	1171	36.996			

Table 8. Comparisons between TIP Form A and Form B mean thresholds by age and sex groups, using Dunn's multiple comparison procedure.

Subject Group	Test M_D in dB	Test C	Test P	Retest M_D in dB	Retest C	Retest P
3 Year						
Male		1.02	N.S.	.89	1.17	N.S.
Female	- .58	.62	N.S.	.83	.89	N.S.
4 Year						
Male	.65	.93	N.S.	.67	.95	N.S.
Female	3.08	4.75	.01	1.84	2.84	.01
5 Year						
Male	1.26	1.87	N.S.	.91	2.77	.05
Female	2.24	2.81	.01	1.87	2.24	.05
6 Year						
Male	.52	.69	N.S.	1.73	3.65	.01
Female	2.81	4.74	.01	2.20	3.71	.01
7 Year						
Male	1.87	3.12	.01	1.98	3.34	.01
Female	1.89	3.14	.01	1.00	1.66	N.S.
8 Year						
Male	1.21	3.20	.01	2.42	2.88	.01
Female	1.90	3.25	.01	1.71	2.92	.01

Note: Plus mean difference (M_D) values indicate Form B mean score larger than Form A mean score; a minus M_D value indicates Form A mean score larger than Form B mean score.

scores, the C values, and probability values for difference between Form means on test and on retest for each age and sex subgroup. For the C computations the value 5.244, which was the test forms F ratio error term, was used in the formula.

Inspection of the pattern of mean differences and probability values shown in Table 8 indicates a trend for the mean threshold on Form B to be more intense than the mean threshold on Form A. The mean of the mean differences between Forms A and B on test was 1.52 decibels, with Form B having the more intense average threshold. On the retest the mean of the mean differences between Forms A and B was 1.50, with Form B having the more intense average threshold.

To test the null hypothesis that there is no difference in test score variance between Form A and Form B of the TIP test, Table 6 was inspected. It is noticeable in Table 6 that all of the standard deviations for age and sex subgroups varied between approximately 3.5 and approximately seven. (Exceptions were Form B test, eight year old male subjects, and retest Form A, three year old females.) When comparing the standard deviations for Form A with the standard deviation for Form B between age and sex subgroups, no comparison gave a ratio over two-to-one. The null hypothesis of no difference in test variance between Form A and Form B could not be rejected, and for the present variability among subject groups for the two test Forms may be assumed to be essentially equivalent.

Comparisons between the Sexes. To test the hypothesis of no difference between males and females within any of the yearly age groups for threshold on TIP Forms A or B, Dunn's multiple comparison procedure (Sparks, 1963) was used. For this computation the error term of 81.139 (error-B in Table 7) was used in the C formula. Table 9 shows the results. None of the C values was significant.

Inspection of the direction of mean differences in Table 9 indicates that in general the females had larger (more intense) thresholds than the males. On the other hand, many of the mean differences (including the negative values in Table 9) were less than one decibel, there were several reversals in direction as a function of age, and there did not appear to be a regular progression towards larger or smaller mean differences as a function of age between males and females. Thus for the present the null hypothesis is retained, and it is concluded for the present that there are not significant sex differences at any age level which need to enter into clinical interpretation of the TIP test results.

To test the null hypothesis of no difference between males and females within any of the age groups with respect to variance of TIP test thresholds the standard deviation values shown in Table 6 were inspected. For this analysis comparisons were made between the standard deviations for the three year old males and the three year old females on test Form A, test Form B, retest Form A, retest Form B, and so forth for each of the yearly age groups. In none of these comparisons was the standard deviation more than twice that of another with which it was compared. The hypothesis of no difference in variability of test scores between the sexes of each age

Table 9. Comparison between males and females for mean TIP test threshold scores within yearly age groups.

Subject Group	Test				Retest			
	Form A		Form B		Form A		Form B	
	M_D	C	M_D	C	M_D	C	M_D	C
3 Year	- .47	.14	.89	.27	-2.92	.87	-2.86	.85
4 Year	- .45	.17	-2.87	1.08	- .78	.29	-1.95	.73
5 Year	.03	.02	-1.05	.34	.94	.35	.10	.03
6 Year	-2.39	1.02	-1.78	2.03	-3.46	1.47	-3.93	1.67
7 Year	-1.50	.64	-1.52	.65	- .47	.18	.55	.23
8 Year	- .75	.32	-1.44	.62	-1.58	.68	- .87	.38

NOTE: A minus sign preceding a mean difference (M_D) value indicates the female mean threshold was larger than the mean threshold for males. No C values significant at .05 level.

group could not be rejected. In these comparisons on some occasions the males had the higher standard deviations, while in others the females had a higher standard deviation. There was no noticeable pattern among Forms or age groups for one sex tending towards larger standard deviations than the other. Thus the test forms can be considered equivalent in variability between the sexes at each of the age groups of interest.

Comparisons among Age Groups. To test the hypothesis of no differences among yearly age groups in TIP thresholds, either for male or for females, Dunn's multiple comparison procedure was followed (Sparks, 1963), using 81.139 as the error term (error term B in Table 7). The three year old group mean TIP test threshold on Form A was compared with the four year old group mean test threshold, the three year old mean test threshold on Form A was compared with the five year old mean threshold, the three year Form A test threshold mean was compared with the six year old mean threshold, and so forth for all the between-age comparisons for test Form A and Form B, and for each retest. Tables 10 and 11 show the results of these C computations for the males and for females. These Tables indicate that in general adjacent age groups were not significantly different from each other, but that distant age groups tended to be significantly different. For example, there were no significant differences between ages three and four or between ages seven and eight in either Table 10 or Table 11. There were frequent significant differences between ages three and eight in both Tables. These observations lead to rejection of the null hypothesis, and to the conclusion

Table 10. C test values for differences among yearly age groups in mean TIP initial test thresholds, males only.

		Ages				
		8 Yr.	7 Yr.	6 Yr.	5 Yr.	4 Yr.
3 Yr.	Form A Test	3.27**	2.58*	1.85	1.34	.08
	Form B Test	3.11**	2.17	1.98	1.17	.12
	Form A Retest	3.50**	2.24*	1.79	.85	-.17
	Form B Retest	2.92**	1.83	1.48	.84	-.10
4 Yr.	Form A Test	3.29**	2.58*	1.85	1.27	
	Form B Test	3.07**	2.10	1.94	1.06	
	Form A Retest	3.81**	2.51*	2.07	1.03	
	Form B Retest	3.13**	2.00	1.66	.94	
5 Yr.	Form A Test	1.94	1.20	.47		
	Form B Test	1.96	.96	.80		
	Form A Retest	2.75*	1.41	.96		
	Form B Retest	2.14	.98	.64		
6 Yr.	Form A Test	1.56	.78			
	Form B Test	1.23	.17			
	Form A Retest	1.89	.47			
	Form B Retest	1.60	.36			
7 Yr.	Form A Test	.78				
	Form B Test	1.07				
	Form A Retest	1.43				
	Form B Retest	1.24				

Note: A minus sign indicates that the mean score of the older group was larger than the mean score for the younger group.

* Significant at .05.
** Significant at .01.

Table 11. C test values for differences among yearly age groups in mean TIP initial test thresholds, females only.

		Ages				
		8 Yr.	7 Yr.	6 Yr.	5 Yr.	4 Yr.
3 Yr.	Form A Test	2.73*	1.88	1.00	1.24	.08
	Form B Test	1.92	1.08	- .10	.40	-1.08
	Form A Retest	3.46**	2.73*	1.39	1.83	.52
	Form B Retest	3.17**	2.67*	.95	1.53	.20
4 Yr.	Form A Test	3.37**	2.27*	1.16	1.38	
	Form B Test	3.85**	1.36	1.27	1.63	
	Form A Retest	3.71**	2.77*	1.09	1.61	
	Form B Retest	3.77**	3.11**	.94	1.60	
5 Yr.	Form A Test	1.43	.52	- .44		
	Form B Test	-1.59	.67	- .60		
	Form A Retest	1.49	.71	- .74		
	Form B Retest	1.55	1.01	- .85		
6 Yr.	Form A Test	2.29*	1.16			
	Form B Test	2.69*	1.55			
	Form A Retest	2.73*	1.76			
	Form B Retest	2.94**	2.27*			
7 Yr.	Form A Test	1.11				
	Form B Test	1.10				
	Form A Retest	.94				
	Form B Retest	.63				

Note: A minus sign indicates that the mean score of the older group was larger than the mean score for the younger group.

* Significant at .05.
** Significant at .01.

that there is a generalized age effect for mean TIP thresholds.

Figures 2 and 3 show the mean thresholds graphically. For these Figures only the initial test data (not retest) for each age group were used. Inspection of these figures gives the impression of change in mean threshold as a function of age, with more acute thresholds for the older age groups. This is so despite the irregularities in the growth curves, especially among the female subjects at age four years and at age six years. (Evident also in both figures is the separation between thresholds obtained on Form A and Form B for both sexes.)

To test the hypothesis of no differences among yearly age groups with respect to variance on TIP thresholds for Form A or Form B, the standard deviation data shown in Table 6 were considered. As with previous inspections of the standard deviation data, from age group to age group no standard deviation approached a ratio of two times that of any other standard deviation in another age group (both for males and for females, for Form A and Form B, on test and on retest). Therefore, the null hypothesis is retained, and it will be assumed there is no difference in variability as a function of age in TIP threshold measurement. The variability across age groups, (standard deviations in Table 6) although of different values at various ages, does not show a consistent trend towards larger or smaller values. Differences in standard deviations appear to be random as a function of age, and a satisfactory best single estimate of standard deviation for TIP thresholds among subject groups would be the statistical mean of the standard deviation values. (According

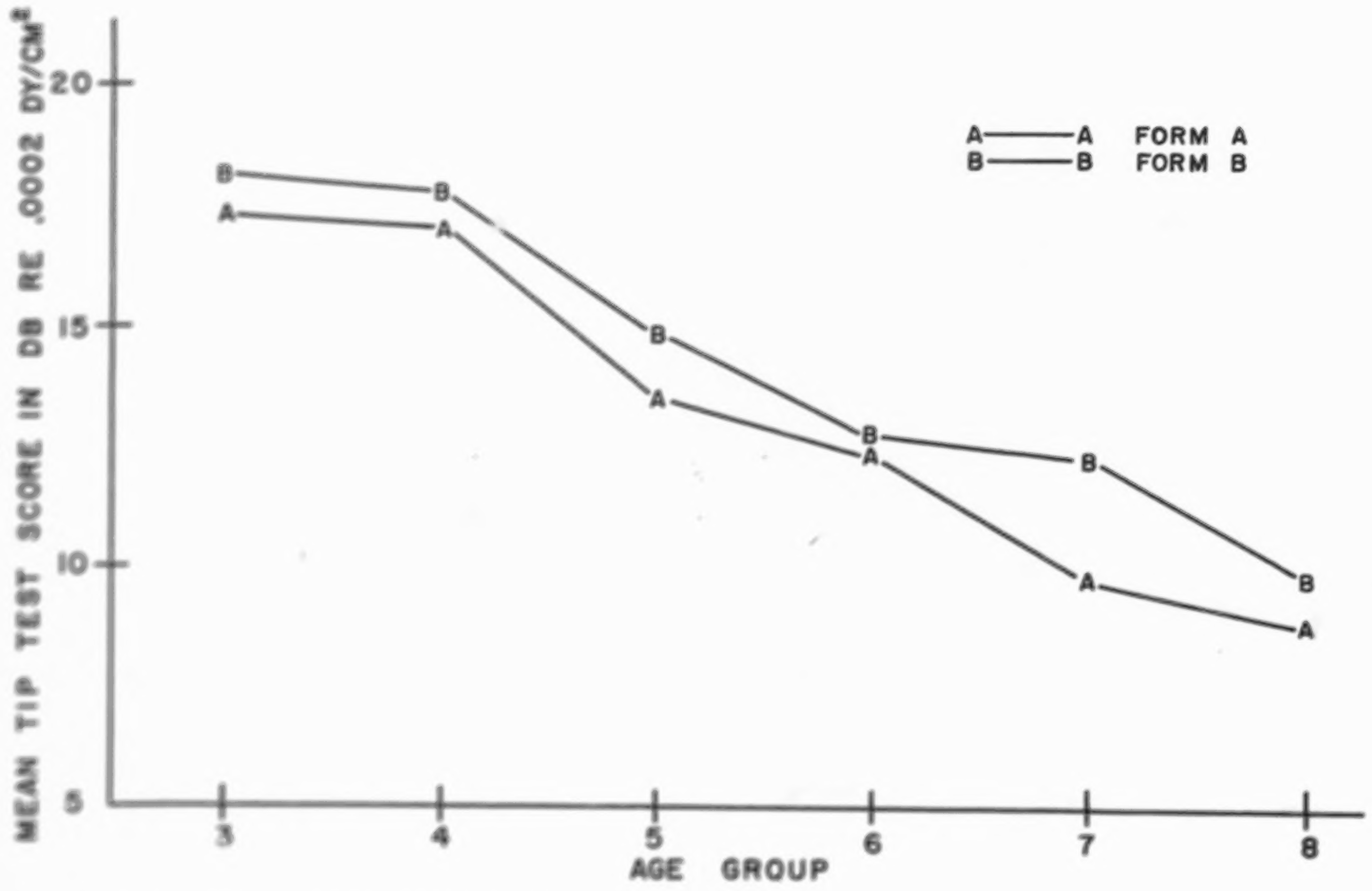


Figure 2. Mean TIP test scores by age group - males.

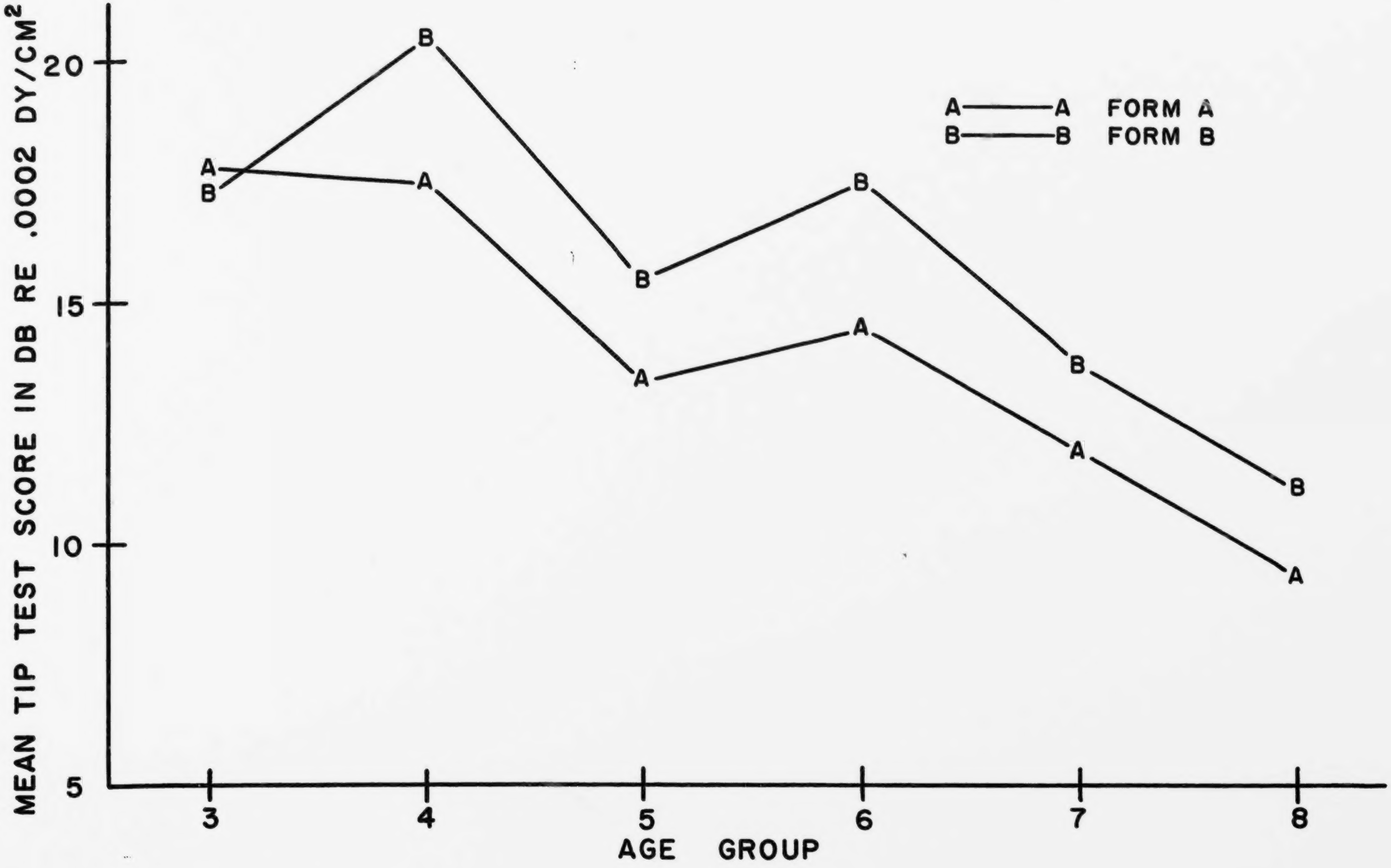


Figure 3. Mean TIP test scores by age group - females.

to an earlier analysis no differences were seen in standard deviations between male and female subjects across the ages studied.) At the bottom of Table 6 are the average standard deviations, computed as means across all subjects for a given Form of TIP on either test or retest.

TIP Threshold Measurement Reliability. Product moment correlations between Form A test and retest thresholds and Form B test and retest thresholds were computed. These correlation values are shown in Table 12. All are significantly different from zero at the .01 level. They are only moderately high, but are satisfactory in view of the restrictions placed upon the data. That is, all subjects were carefully preselected to have normal hearing acuity (thresholds for pure tones no worse than 10 decibels, ASA, 1951 standard). This severe restriction in range of hearing acuity of necessity resulted in a restricted range of speech reception test scores.

A more satisfactory method of considering test-retest reliability is the absolute decibel difference between test and retest thresholds. To approach the problem in this manner, the difference in decibels between initial test threshold on each Form of TIP and retest threshold on each Form of TIP was calculated for each child. The range of these differences and the mean difference for each age and sex group are shown in Table 13. These differences were computed without regard to direction of difference. (If direction of difference were considered, the tendency would be to underestimate the test-retest difference due to algebraic cancellation of differences.)

Table 13 indicates that although there were a few cases of

Table 12. Product-moment correlations between TIP initial test and retest thresholds

	Form A Test	Form A Retest	Form B Test
Form A Retest	0.731		
Form B Test	0.859		
Form B Retest		0.860	0.725

Note: All r values significantly different from zero at the .01 level.

Table 13. Range of differences, and absolute mean difference in decibels between test and retest for TIP thresholds.

Subject Group	TIP Form A		TIP Form B	
	Range	Mean	Range	Mean
3 Year				
Male	0-7	3.6	0-11	3.8
Female	0-5	2.4	0-13	3.9
4 Year				
Male	0-8	3.8	0-10	3.4
Female	1-11	3.5	0-11	3.7
5 Year				
Male	0-7	2.6	0-11	3.5
Female	0-7	2.5	0-8	3.9
6 Year				
Male	0-11	3.3	0-10	3.2
Female	0-10	3.3	0-11	3.1
7 Year				
Male	0-9	3.1	0-9	2.4
Female	0-12	2.8	0-16	3.4
8 Year				
Male	0-10	4.3	0-10	3.8
Female	0-10	3.3	0-8	3.0
Over-all Mean		3.2		3.4
Sample Std. Dev.		2.76		3.06
Est. Population Std. Dev.		2.77		3.07

relatively large or small ranges, there did not appear to be a systematic age or sex effect in this respect. The mean difference scores were relatively homogenous, without an evident age or sex effect. The overall-group means are approximately equal. The standard deviations of the distribution of differences between test and retest thresholds across all 295 subjects were computed. The sample standard deviations are indicated at the bottom of Table 13, as are estimated population standard deviations. These values, being close to three decibels, provide our best single estimate of the standard error of estimate of test-retest for TIP Form A and Form B. For the present the estimate of test-retest reliability (the range within which two-thirds of the test-retest differences would be expected to fall) is ±3dB.

TIP Intelligibility Curve. Preliminary inspection of the raw data on the TIP test for subjects across age groups indicated that there was stability of articulation curve (items correct as a function of dB presentation level) for the sexes, and separate intelligibility curves were not considered for males and females.

An intelligibility curve was drawn for each age group for each TIP Form by tabulating total number of words presented at each dB level at which words were given. The per cent correct responses at each presentation level was derived, and used to draw an intelligibility curve for that age group. The slopes of the curves (per cent change in intelligibility per dB change in presentation level) between twenty and eighty per cent were as follows:

		Form A	Form B	Mean
3 year group,	per cent per dB	3.93	5.71	4.82
4 year group,	per cent per dB.	4.28	4.36	4.32
5 year group,	per cent per dB:	4.29	4.44	4.36
6 year group,	per cent per dB:	5.57	4.90	5.24
7 year group,	per cent per dB:	5.00	4.04	4.52
8 year group,	per cent per dB:	4.62	4.53	4.58
	Means:	4.62	4.66	4.64

The homogeneity of these slopes, lack of systematic change in slope as a function of age, and similarity of the terminal points of the curves as they were drawn graphically lead to an articulation curve of best fit over all subject age groups.

The intelligibility curve for each age group was adjusted so that 50 per cent was given the value of zero dB, and the data points for each age group were noted as deviations from this referent. The data points for all age groups were plotted together, grouped within each five dB step of intensity change, and a curve was drawn to represent the over-all curve of best fit. This curve could reasonably be approximated by a straight line over its middle section (although it showed curvature characteristic of an ogive curve). Its upper end did not change as rapidly as its lower end, reflecting the occasional lack of one hundred per cent responses on the TIP materials by the younger subjects. However, using a less steep slope for the upper per cent areas than necessary for older subjects did not necessitate changing the slope of the curve over its critical area, namely the middle range. The composite intelligibility function for the TIP Forms A and B may be described as follows:

> change from twelve per cent to 75 per cent intelligibility over a range of 14.5 dB (4.3 per cent intelligibility change per dB intensity change)

zero per cent intelligibility at twelve dB below decibel level which gives twelve per cent intelligibility

ninety per cent intelligibility at five dB above decibel level which gives 75 per cent intelligibility

one hundred per cent intelligibility at fifteen dB above decibel level which gives 75 per cent intelligibility

This curve is shown on Figure 1 and on the TIP SCORING SHEET in Appendix A. It may be generated in template form to fit the clinical forms in use in an audiology program as an aid to drawing the curve of best fit for a given child's test responses, or it may be hand drawn to produce a line of best fit for the response data on a child.

Analysis Related to DIP Test

The reader will recall that the DIP test was administered at each of three intensity levels: SRT + 0 dB, SRT + 5 dB, and SRT + 10 dB, and that it was scored for 48 items. Tables 14 and 15 show the mean DIP scores obtained by male and female subjects at each age level on initial testing and on retesting. Because of the rotational orders utilized, Form 3 of DIP was never given at Level SRT + 10 dB.

As the initial analysis of the DIP data, three Lindquist Type III analyses of variance were done (Lindquist, 1953), one for each presentation level. Tables 16, 17, and 18 are summaries of these analyses.

Comparisons among DIP Forms. Noticeable in all three of Tables 16, 17, and 18 are the nonsignificant F ratios for DIP Forms. Therefore the null hypothesis of no difference in mean score between DIP Forms 1, 2, and 3 was retained, and three forms of the DIP test

Table 14. Mean DIP scores for age and sex subgroups on the initial test at each presentation level (dB above SRT obtained on TIP test).

Subject Group	Form 1			Form 2			Form 3	
	SRT + 0	SRT + 5	SRT + 10	SRT + 0	SRT + 5	SRT + 10	SRT +0	SRT + 5
3 Year								
Males	12.50	21.50	31.67	11.25	22.50	33.67	11.33	23.75
Females	21.25	17.00	34.40	16.00	24.33	32.29	16.83	30.33
4 Year								
Males	21.00	30.71	38.00	19.33	28.33	41.69	20.38	30.11
Females	23.17	29.17	38.31	21.00	22.33	37.33	22.84	32.15
5 Year								
Males	19.00	33.00	40.57	23.63	30.50	39.89	21.09	30.28
Females	18.75	28.33	40.74	27.00	29.57	39.57	21.70	33.20
6 Year								
Males	19.50	29.27	41.93	26.22	33.00	41.00	22.77	34.61
Females	19.50	32.00	41.87	23.40	38.20	40.13	23.83	32.22
7 Year								
Males	19.78	31.22	40.46	20.00	34.25	40.56	24.16	32.69
Females	29.14	35.33	40.62	22.11	41.00	43.19	24.85	35.00
8 Year								
Males	22.50	36.29	41.24	27.29	38.80	43.25	27.00	35.22
Females	22.50	31.00	42.29	27.50	36.92	42.91	23.22	35.07

Table 15. Mean DIP scores for age and sex subgroups on the retest at each presentation level (dB above SRT obtained on TIP retest)

Subject Group	Form 1			Form 2			Form 3	
	SRT + 0	SRT + 5	SRT + 10	SRT + 0	SRT + 5	SRT + 10	SRT + 0	SRT + 5
3 Year								
Males	14.38	25.00	23.67	18.50	21.00	36.42	13.83	24.92
Females	25.00	19.67	34.20	19.50	19.33	34.00	14.17	30.00
4 Year								
Males	16.83	32.29	38.44	27.00	27.67	39.23	18.92	30.66
Females	30.67	32.50	37.31	16.57	26.67	38.42	21.25	29.15
5 Year								
Males	21.33	30.17	39.65	23.00	29.83	37.89	21.25	29.37
Females	15.75	26.00	37.63	27.00	27.86	41.57	22.30	35.40
6 Year								
Males	18.25	29.64	43.46	22.00	34.50	42.00	21.52	32.85
Females	19.50	33.57	41.80	17.50	33.80	43.47	25.92	31.39
7 Year								
Males	18.00	30.78	43.46	22.00	34.50	41.65	23.39	31.56
Females	27.29	32.78	41.61	22.44	39.25	45.06	23.23	34.12
8 Year								
Males	16.50	34.36	40.16	29.71	33.40	42.44	24.15	32.77
Females	26.17	31.20	39.14	27.13	33.00	44.91	20.95	34.93

Table 16. Summary of Lindquist Type III analysis of variance for DIP at level SRT + 0 dB, Forms 1, 2, 3.

Source	Sum of Sq. s	d. f.	Mean Sq.	Error Term	F	P
Between Subjects	26678.440	292	91.364			
Age-Sex	4582.580	11	416.598	Error (B)	5.500	.01
Forms	296.220	2	148.110	Error (B)	1.955	N.S.
Age-Sex x Forms	2333.960	22	106.089	Error (B)	1.400	N.S.
Error (B)	19465.680	257	75.742			
Within Subjects	10231.500	293	34.920			
Test-Retest	45.330	1	45.330	Error (W)	1.307	N.S.
Age-Sex x Test-Retest	225.900	11	20.536	Error (W)	0.5919	N.S.
Forms x Test-Retest	41.880	2	20.940	Error (W)	0.603	N.S.
Forms x Age-Sex x Test-Retest	1002.340	22	45.561	Error (W)	1.313	N.S.
Error (W)	8916.05	257	34.693			
Total	30909.940	585	63.094			

Table 17. Summary of Lindquist Type III analysis of variance for DIP at SRT + 5 dB, Forms 1, 2, and 3.

Source	Sum of Sq. s	d. f.	Mean Sq.	Error Term	F	P
Between Subjects	25116. 740	292	86. 016			
Age-Sex	5718. 370	11	519. 852	Error B	7. 731	. 01
Forms	59. 160	2	29. 580	Error B	0. 440	N. S.
Age-Sex x Forms	2058. 950	22	93. 589	Error B	1. 391	N. S.
Error (B)	17280. 260	257	67. 238			
Within Subjects	8911. 000	293	30. 413			
Test-Retest	120. 750	1	120. 750	Error (W)	3. 781	. 05
Age-Sex x Test-Retest	170. 570	11	15. 506	Error (W)	0. 486	N. S.
Forms x Test-Retest	44. 960	2	22. 480	Error (W)	0. 704	N. S.
Forms x Age-Sex x Test Retest	366. 880	22	16. 676	Error (W)	0. 522	N. S.
Error (W)	8207. 840	257	31. 937			
Total	34027. 740	585	58. 167			

Table 18. Summary of Lindquist Type III analysis of variance for DIP at level SRT + 10 dB, Forms 1, 2, and 3.

Source	Sum of Sq. s	d. f.	Mean Sq.	Error Term	F	P
Between Subjects	14620.740	292	50.071			
Age-Sex	4554.120	11	414.010	Error B	12.139	.01
Forms	106.870	1	106.870	Error B	3.133	N. S.
Age-Sex x Form	784.870	11	71.352	Error B	2.092	N. S.
Error (B)	9174.880	269	34.107			
Within Subjects	5855.500	293	19.985			
Test-Re-Test	0.010	1	0.010	Error (W)	0.000	
Age-Sex x Test-Retest	197.790	11	17.980	Error (W)	0.933	N. S.
Forms x Test-Retest	129.180	1	129.180	Error (W)	6.702	.05
Forms x Age-Sex x Test-Retest	343.430	11	31.221	Error (W)	1.620	N. S.
Error (W)	5185.090	269	19.275			N. S.
Total	20476.240	585	35.002			N. S.

will be accepted as essentially equivalent with respect to test score obtained. For subsequent analyses subject responses to the three Forms usually were pooled.

To test the null hypothesis of no difference in DIP score variance among Forms 1, 2 and 3 at each presentation level, Tables 19 and 20 were inspected. These Tables indicate the standard deviations of distributions of DIP scores for age and sex subgroups at each presentation level. The analysis compared the standard deviations at respective presentation levels for sex and age subgroups, across DIP Forms. That is, the standard deviation on initial test (Table 19) of 6.16 for three year old males listening to Form 1 at level SRT + 0 dB was compared with the standard deviation of 6.65 for three year old males listening to Form 2 at level SRT + 0 dB, and with the standard deviation of 8.62 for Form 3 at level SRT + 0 dB. In no case was the ratio between a pair of standard deviations greater than approximately two-to-one for three year old males, for four year old males and females, or for eight year old males.

The large majority of the standard deviations ranged from approximately three to approximately eight. A scattered group, indicated by asterisks in Table 19 were relatively large or relative small in comparison to the general level of the values. (In one case, standard deviation for five year old females, DIP Form 2 at level SRT + 0, only one subject took the test and therefore there was no variability.) These scattered deviant values account for some ratios between standard deviations exceeding two to one. For example, the standard deviation of 3.61 for five year old males, Form 1 at SRT + 0 dB

Table 19. Standard deviation of distribution of DIP scores for age and sex groups at each presentation level, initial test.

Subject Group	Form 1			Form 2			Form 3	
	SRT + 0	SRT + 5	SRT + 10	SRT + 0	SRT + 5	SRT + 10	SRT + 0	SRT + 5
3 Year								
Males	6.16	5.74	7.21	6.65	4.95	5.38	8.62	6.09
Females	6.60	8.19	4.29	7.07	13.05	5.82	*1.73	2.83
4 Year								
Males	6.99	8.16	5.49	*11.85	4.07	6.57	9.42	8.57
Females	4.71	7.14	3.82	7.09	7.97	5.54	5.06	6.67
5 Year								
Males	3.61	4.65	3.90	*9.40	3.83	5.11	4.42	*8.99
Females	*12.28	5.51	3.64	*0.00 (N 1)	3.74	3.04	6.85	8.74
6 Year								
Males	*8.81	7.84	3.27	5.61	7.67	3.62	4.68	7.62
Females	*11.92	7.44	3.39	6.82	*1.30	4.06	7.72	8.01
7 Year								
Males	5.45	2.99	3.53	*10.20	7.14	6.43	6.06	8.46
Females	5.64	4.15	3.24	6.53	5.35	*2.08	4.89	6.24
8 Year								
Males	7.78	8.87	4.20	4.35	5.45	*2.59	8.03	4.11
Females	4.37	6.86	*2.78	5.07	6.29	*2.47	5.08	5.37

Table 20. Standard deviations of distribution of DIP scores for age and sex groups at each presentation level, retest.

Subject Group	Form 1			Form 2			Form 3	
	SRT + 0	SRT + 5	SRT + 10	SRT + 0	SRT + 5	SRT + 10	SRT + 0	SRT + 5
3 Year								
Males	6.30	8.52	12.11	8.27	2.83	7.46	5.88	4.96
Females	2.94	7.51	6.35	4.95	14.74	11.60	4.85	4.57
4 Year								
Males	4.88	8.10	4.77	8.19	6.53	3.94	9.40	5.53
Females	7.31	4.85	7.30	6.55	10.39	5.26	4.85	8.30
5 Year								
Males	7.51	6.40	6.11	5.86	8.21	5.24	7.61	6.58
Females	10.63	3.61	5.32	0.00 (N 1)	6.82	5.04	5.71	7.14
6 Year								
Males	5.32	8.48	3.38	7.65	2.74	3.12	6.96	5.86
Females	10.09	5.68	3.82	8.54	3.96	3.47	7.51	6.17
7 Year								
Males	8.00	7.87	2.28	6.69	7.59	4.80	5.46	7.65
Females	9.66	6.22	4.68	8.25	3.20	1.91	6.46	6.88
8 Year								
Males	2.12	6.31	4.85	9.72	5.98	3.45	9.89	4.13
Females	5.19	11.26	7.65	6.58	8.92	1.95	9.08	4.74

compared with the same group's standard deviation of 9.40 for DIP Form 2 at SRT + 0 dB, gave well over a two-to-one ratio. In other instances the ratios between standard deviations to be compared approached three-to-one. However, because of the great degree of scattering of large and small standard deviations throughout the Table and lack of noticeable pattern of such values, it is concluded that the ratios greater than two-to-one between standard deviations were due to sampling errors, and that DIP Forms 1, 2 and 3 produce essentially the same distribution of test scores. (The same conclusion was reached by a similar analysis of Table 20.)

Comparisons between the Sexes. To test the null hypothesis of no difference between male and female subjects in DIP scores at each presentation level, a series of Dunn's multiple comparison procedural tests was done (Sparks, 1963). The mean score by three year old males across all three Forms of the DIP combined, was compared with the mean DIP score across all three Forms combined for three year old females. For these comparisons the presentation levels (across the three Forms) were kept separate. The mean value for each sex within each age level, for each presentation level is shown in Table 21. Included in Table 21 are the numbers of children (N) in each age-by-sex category for each level. (These numbers are not repeated for the retest because the same children were retested as were tested.) Table 22 shows the mean differences between scores and the computed C values. No C value in Table 22 reaches the .05 level of significance. An inspection of the direction of mean differences throughout Table 22 does not indicate a consistent trend for males to have

Table 21. Mean DIP test scores combined across Forms 1, 2 and 3 for males and for females at each age group, and at each presentation level.

			Test					Retest	
Subject	SRT		SRT		SRT		SRT	SRT	SRT
Group	+0 dB	N	+5 dB	N	+10 dB	N	+0 dB	+5 dB	+10 dB
3 Year									
Males	11.69	18	22.58	18	32.67	18	15.58	23.64	30.04
Females	18.03	12	23.89	12	33.35	12	19.56	23.00	34.10
4 Year									
Males	20.24	22	29.72	22	39.85	22	20.92	30.21	38.84
Females	22.34	25	27.88	25	37.82	25	22.83	29.44	37.87
5 Year									
Males	21.24	23	28.20	23	40.23	23	21.86	29.79	38.77
Females	22.48	15	30.37	15	40.18	15	21.68	29.75	39.60
6 Year									
Males	22.83	30	32.29	30	41.47	30	20.59	32.33	42.73
Females	22.24	30	34.14	30	41.00	30	20.97	32.92	42.64
7 Year									
Males	21.31	29	32.72	29	40.51	29	21.13	32.28	42.56
Females	25.37	29	37.11	29	41.91	29	24.32	35.38	43.34
8 Year									
Males	25.60	28	36.77	28	42.25	28	23.45	33.51	41.30
Females	24.41	32	34.33	32	42.60	32	24.75	33.04	42.03

Table 22. Comparison between males and females within yearly age groups for DIP Test at levels SRT + 0 dB, + 5 dB, and + 10 dB, using Dunn's multiple comparison procedure.

Subject	Test						Retest					
	SRT + 0		SRT + 5		SRT + 10		SRT + 0		SRT + 5		SRT + 10	
Group	M_D	C	M_D	C	M_D	C	M_D	C	M_D	C	M_D	C
3 Year	+6.34	1.99	+1.31	.44	+ .68	.31	+3.88	1.16	- .64	.21	+4.06	1.86
4 Year	+2.10	.83	-1.84	.77	-2.03	1.19	+1.91	.75	- .77	.32	- .97	.57
5 Year	+1.24	.43	+2.17	.80	- .07	.04	- .18	.06	- .04	.01	+ .83	.43
6 Year	- .59	.26	+1.85	.87	- .47	.31	- .16	.07	+ .59	.28	- .09	.06
7 Year	+4.06	1.77	+4.39	2.04	+1.40	.92	+3.19	1.39	+3.10	1.44	+ .78	.51
8 Year	-1.19	.53	-2.44	1.15	+ .35	.23	+1.30	.58	- .47	.22	+ .73	.48

Note: A plus mean difference (M_D) value indicates female mean score larger than male mean score. A minus M_D value indicates male mean larger than the female mean. No C values significant at .05 level.

Error term for C at SRT + 0 dB 75.742 (Error B in Table 16).

Error term for C at SRT + 5 dB 67.238 (Error B in Table 17).

Error term for C at SRT + 10 dB 34.107 (Error B in Table 18).

larger scores than females, nor a trend as a function of age for direction of DIP score difference at any presentation level. The null hypothesis of no difference between males and females was retained, and for the present it will be assumed that boys and girls, in the age range studied, have essentially the same DIP scores for a given presentation level.

To test the hypothesis of no difference between males and females within any of the yearly age groups three to eight inclusive in variance of DIP scores at each presentation level, the standard deviations shown in Table 19 were inspected. In this instance, the male standard deviation was compared with the female deviation for each Form and each presentation level. The overall finding from Table 19 is that this comparison did not indicate standard deviation ratios of more than two-to-one. Notable exceptions were instances such as three year olds, Form 2, level SRT + 5 dB; three year olds, Form 3, level SRT + 0 dB, five year olds, Form 1, level SRT + 0 dB. However, there appeared to be no pattern, either for age or sex groups having larger standard deviations, nor for a consistent trend according to form or presentation level. Therefore, the data fail to reject the null hypothesis of no difference in variance of DIP scores between the sexes. It is assumed that overall there is no significant difference in standard deviation distribution of DIP test scores between the sexes (assuming the "atypical" ratios in Table 19 were due to sampling errors). This conclusion is strengthened by comparing Table 20 with Table 19. Many of the comparisons where large ratios were found on initial test had reduced ratios on retest. For example, in Table 20 the five year old children on retest of Form 1, level SRT + 0 dB, had a small ratio in standard deviations.

Comparison among Age Groups. To test the null hypothesis of no difference in DIP scores at any of the presentation levels among yearly age groups three to eight inclusive, Dunn's multiple comparison procedure C tests were done (Sparks, 1963). These were computed separately for males and for females at each presentation level, but by pooling data across DIP Forms. Only initial test scores were used. Tables 23, 24 and 25 show the C values for males, probability values of the C's, and the differences between means. The mean DIP test scores for each age group is shown along the margins of the Tables. Tables 26, 27 and 28 show similar data and C values for the female subjects.

This series of tables shows a number of instances of statistically significant differences between mean DIP test scores between age groups. Especially noticeable in Tables 23, 24 and 25 are the significant differences between three year olds and all of the older age groups, but only occasional significant differences among other age groups. In Tables 26, 27 and 28, for female subjects, there also are statistically significant differences noticeable between the younger and older age groups, but not frequent statistically differences between adjacent age groups. On the basis of these C values and their associated probabilities, the hypothesis of there being no differences in mean DIP scores across ages at any of the presentational levels is rejected.

Table 23. Dunn's C test values for differences among yearly age groups in mean DIP test scores, at level SRT + 0 dB, males only.

Subject Group	8 Year (25.60)	7 Year (21.31)	6 Year (22.83)	5 Year (21.24)	4 Year (20.24)
3 Year (11.69)	5.30** (M_D 13.91)	3.69** (M_D 9.62)	4.40** (M_D 11.14)	3.66** (M_D 9.55)	3.08** (M_D 8.55)
4 Year (20.24)	2.16 (M_D 5.36)	43 (M_D 1.07)	1.06 (M_D 2.59)	.39 (M_D 1.00)	
5 Year (21.24)	1.77 (M_D 4.36)	.03 (M_D .07)	.66 (M_D 1.59)		
6 Year (22.83)	1.21 (M_D 2.77)	.64 (M_D 1.45)			
7 Year (21.31)	1.88 (M_D 4.29)				

Note: Value in parentheses beneath each age group is mean DIP test score for age group at SRT + 0 dB.

Difference between means (M_D) of age groups are indicated below each C value.

* Significant at .05.
** Significatn at .01.

Error term for C at SRT + 0 dB 75.742 (Error B in Table 16).

Table 24. Dunn's C test values for differences among yearly age groups in mean DIP test scores, at level SRT + 5 dB, males only.

Subject Group	8 Year (36.77)	7 Year (32.72)	6 Year (32.29)	5 Year (28.20)	4 Year (29.72)
3 Year (22.58)	5.74** (M_D 14.19)	4.12** (M_D 10.14)	3.97** (M_D 9.71)	2.28** (M_D 5.62)	2.69* (M_D 7.04)
4 Year (29.70)	3.02** (M_D 7.05)	1.29 (M_D 3.00)	1.12 (M_D 2.57)	.62 (M_D 1.52)	
5 Year (28.20)	3.69** (M_D 8.57)	1.97 (M_D 4.52)	1.79 (M_D 4.09)		
6 Year (32.29)	2.08 (M_D 4.48)	4.09** (M_D 10.43)			
7 Year (32.72)	1.87 (M_D 4.05				

Note: Value in parenthesis beneath each age group is mean DIP test score for that age group at SRT + 5 dB.

Differences between means (M_D) of age groups are indicated below C values.

* Significant at .05.

** Significant at .01.

Error term for C at SRT + 5 dB 67.238 (Error B in Table 17).

Table 25. Dunn's C test values for differences among yearly age groups in mean DIP test scores, at level SRT + 10 dB, males only.

Subject Groups	8 Year (42.25)	7 Year (40.51)	6 Year (41.47)	5 Year (40.23)	4 Year (39.85)
3 Year (32.67)	5.44** (M_D 9.58)	4.47** (M_D 7.84)	5.05** (M_D 8.80)	4.32** (M_D 7.56)	3.87** (M_D 7.18)
4 Year (39.85)	1.45 (M_D 2.40)	.10 (M_D .66)	.99 (M_D 1.62)	.22 (M_D .38)	
5 Year (40.23)	1.22 (M_D 2.02)	.17 (M_D .28)	.77 (M_D 1.24)		
6 Year (41.47)	.51 (M_D .78)	.63 (M_D .96)			
7 Year (40.51)	1.12 (M_D 1.74)				

Note: Value in parentheses beneath each age group is mean DIP test score for that age group at SRT + 10 dB.

Differences between means (M_D) of age groups are indicated below C values.

* Significant at .05.
** Significant at .01.

Error term for C at SRT + 10 dB 34.107 (Error B in Table 18).

Table 26. Dunn's C test values for differences among yearly age groups in mean DIP test scores, at level SRT + 0 dB, females only.

Subject Groups	8 Year (24.41)	7 Year (25.37)	6 Year (22.24)	5 Year (22.48)	4 Year (22.34)
3 Year (18.03)	2.16 (M_D 6.38)	2.45* (M_D 7.34)	1.39 (M_D 4.11)	1.32 (M_D 4.45)	1.41 (M_D 4.31)
4 Year (22.34)	.89 (M_D 2.07)	1.28 (M_D 3.03)	.04 (M_D .10)	.05 (M_D .14)	
5 Year (22.48)	.03 (M_D .07)	1.11 (M_D 2.89)	.09 (M_D .24)		
6 Year (22.24)	.98 (M_D 2.17)	1.38 (M_D 3.13)			
7 Year (25.37)	.43 (M_D .96)				

Note: Value in parentheses beneath each age group is mean DIP test score for that age group at SRT + 0 dB.

Differences between means (M_D) of age groups are indicated below each C value.

* Significant at .05.

** Significant at .01.

Error terms for C at SRT + 0 dB 75.742 (Error B in Table 16).

Table 27. Dunn's C test values for differences among yearly age groups in mean DIP test scores, at level SRT + 5 dB, females only.

Subject Groups	8 Year (34.33)	7 Year (37.11)	6 Year (34.14)	5 Year (30.37)	4 Year (27.88)
3 Year (23.89)	3.76* (M_D 10.44)	4.69** (M_D 13.22)	3.65** (M_D 10.25)	2.04 (M_D 6.48)	1.39 (M_D 3.99)
4 Year (27.88)	2.96** (M_D 6.45	4.14** (M_D 9.23)	2.82* (M_D 6.26)	.93 (M_D 2.49)	
5 Year (30.37)	1.54 (M_D 3.96)	2.58* (M_D 6.74)	1.46 (M_D 3.77)		
6 Year (34.14)	.09 (M_D .19)	1.39 (M_D 2.97)			
7 Year (37.11)	1.32 (M_D 2.78)				

Note: Value in parentheses beneath each age group is mean DIP test score for that age group at SRT + 5 dB.

Differences between means (M_D) of age groups are indicated below each C value.

* Significant at .05.

** Significant at .01.

Error term for C at SRT + 5 dB 67.238 (Error B in Table 17).

Table 28. Dunn's C test values for differences among yearly age groups in mean DIP test scores, at level SRT + 10 dB, females only.

Subject Groups	8 Year (42.60)	7 Year (41.91)	6 Year (41.00)	5 Year (40.16)	4 Year (37.82)
3 Year (33.35)	4.67** (M_D 9.25)	4.27** (M_D 8.56)	3.83** (M_D 7.65)	3.01** (M_D 6.81)	2.18 (M_D 4.77)
4 Year (37.82)	3.06** (M_D 4.78)	2.57* (M_D 4.09)	2.01 (M_D 3.18)	1.23 (M_D 2.34)	
5 Year (40.16)	1.33 (M_D 2.44)	.95 (M_D 1.75)	.45 (M_D .84)		
6 Year (41.00)	1.05 (M_D 1.60)	.60 (M_D .91)			
7 Year (41.91)	.46 (M_D .69)				

Note: Value in parentheses beneath each age group is mean DIP test score for that age group at SRT + 10 dB.

Differences between means (M_D) of age groups are indicated below each C value.

* Significant at .05.

** Significant at .01.

Error term for C at SRT + 10 dB 34.107 (Error B in Table 18).

Using the mean DIP test scores shown along the margins of Tables 23 through 28, Figures 4 and 5 were prepared for male and female subjects respectively. These Figures make apparent the general growth trends across ages. In most instances there are especially noticeable changes in DIP scores for the groups three to four years of age, with less rapid change in score beyond that. (An exception to this trend is seen in Figure 5, female subjects at presentation level SRT + 5 dB). The apparent decline in mean DIP test score for this group of subjects at SRT + 5 dB for the eight year old group, may be because there was a spuriously high set of test scores for the female subjects at SRT + 5 dB at ages six and seven years of age, giving the appearance of sustained and rapid growth over the middle ages. Such a possibility could be ascribed to sampling errors.

DIP Test Reliability. As an initial measure of test-retest reliability, product moment correlations at each presentational level between initial testing and retesting on DIP for all subjects combined were computed. Table 29 shows these correlations. All of them are significantly different from zero at the .01 level.

As a more meaningful clinical measure of DIP reliability the difference between the initial test score and the retest score on DIP Forms was obtained. In this computation the direction of the difference was ignored, and the mean absolute difference in test score was computed for age and sex sub-groups. Table 30 shows the range of differences, mean differences and standard deviations of the distributions of absolute differences. According to this Table the mean and range of test to retest differences across subjects and ages was

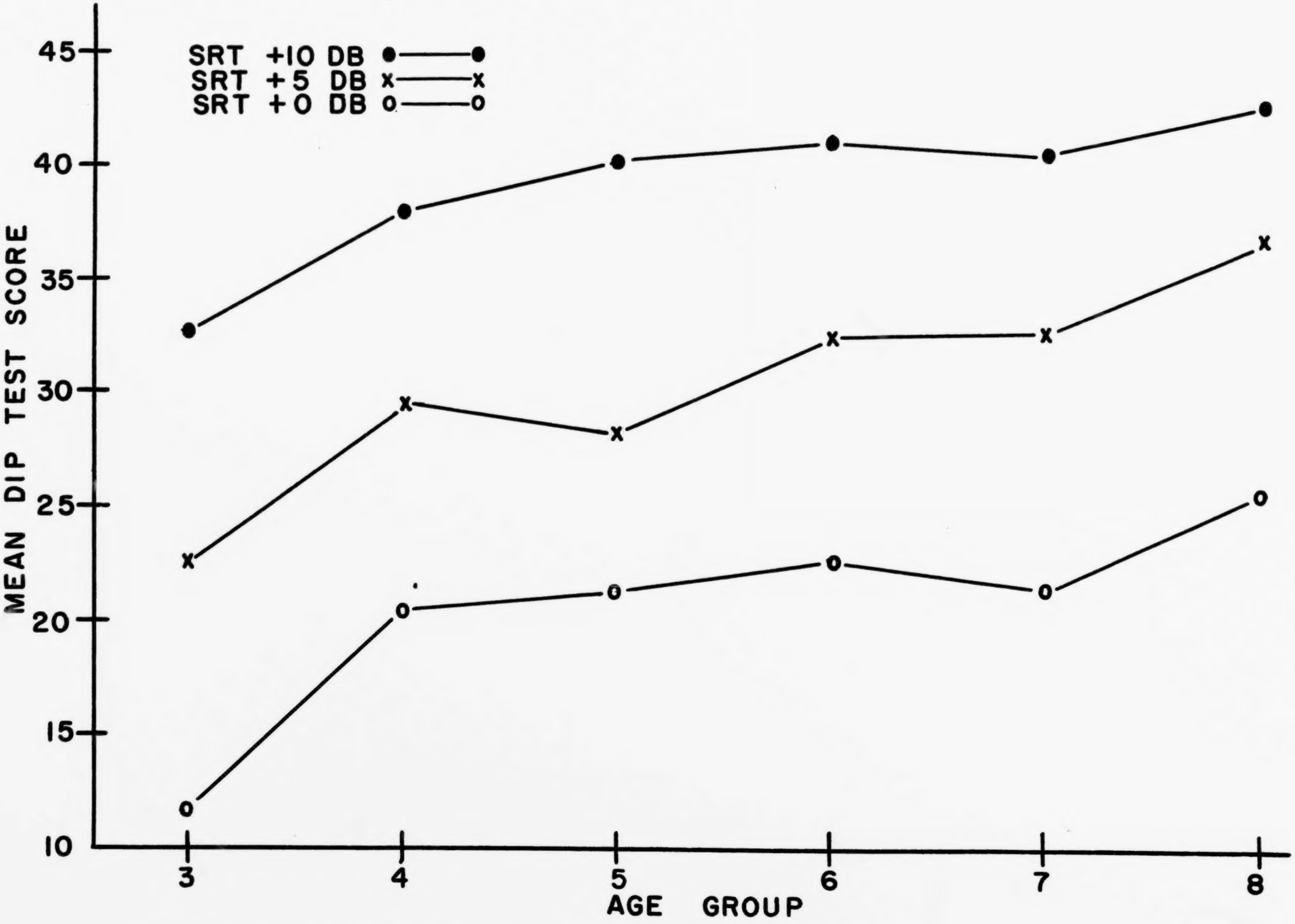


Figure 4. Mean DIP test scores by age group - males.

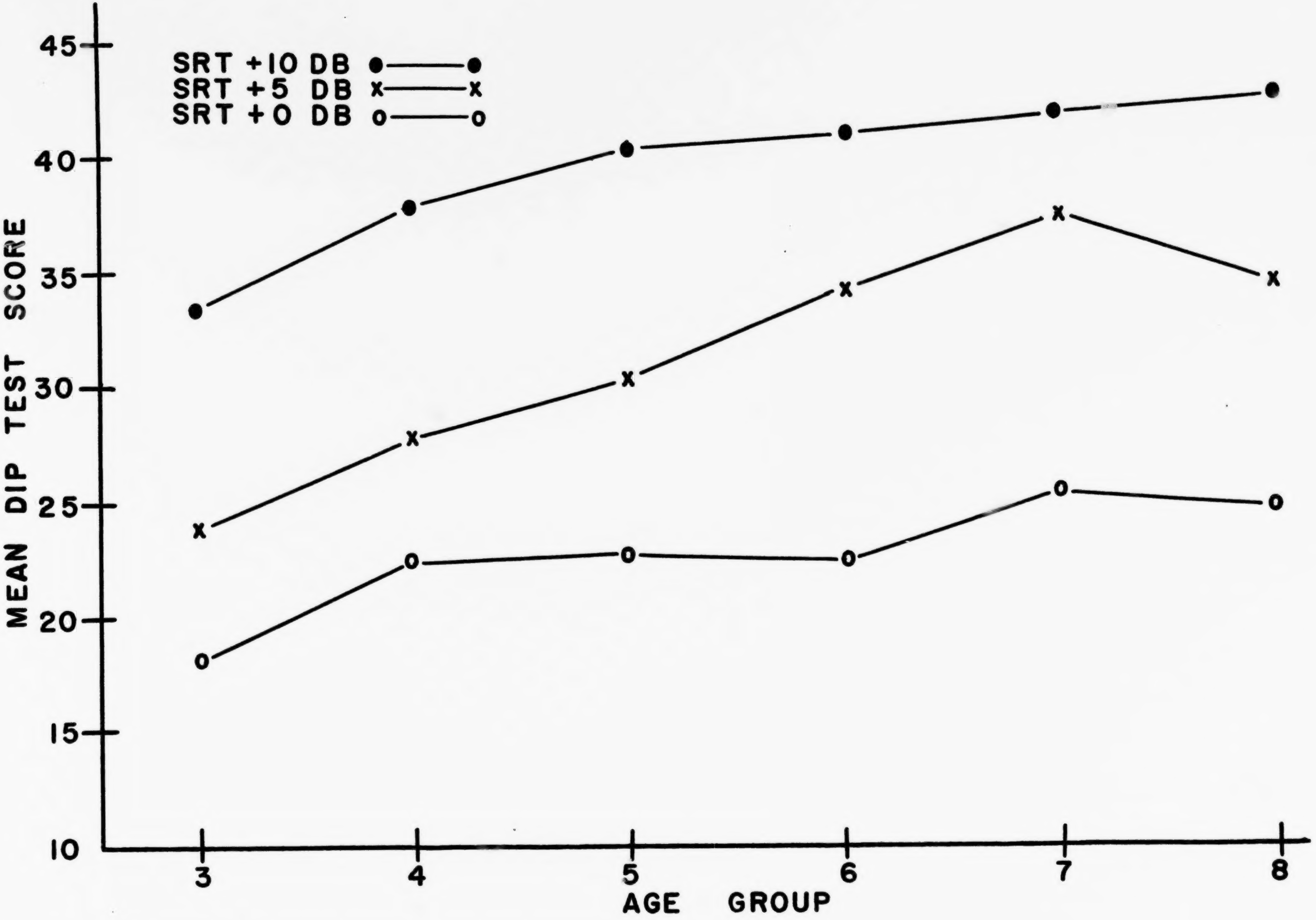


Figure 5. Mean DIP test scores by age group - females.

Table 29. Product moment correlations at each presentation level between initial test and retest scores on DIP for males and females combined (N 295).

Initial Test	Retest						
	Form 1 at SRT			Form 2 at SRT			Form 3 at SRT
	+0	+5	+10	+0	+5	+10	+0
Form 1 at SRT + 0 dB	.59						
Form 1 at SRT + 5 dB		.42					
Form 1 at SRT + 10 dB			.45				
Form 2 at SRT + 0 dB				.36			
Form 2 at SRT + 5 dB					.59		
Form 2 at SRT + 10 dB						.53	
Form 3 at SRT + 0 dB							.42

Note: All r values significant at .01.
Form 3 not given at SRT + 10 dB.

Table 30. Range of differences and absolute mean differences between test and retest for DIP scores.

Subject Group	Level SRT + 0 dB			Level SRT + 5 dB			Level SRT + 10 dB		
	Range	Mean Diff.	Std. Dev.	Range	Mean Diff.	Std. Dev.	Range	Mean Diff.	Std. Dev.
3 Year			4.43			2.91			6.79
Male	1-18	6.17		0-7	3.83		2-35	6.67	
Female	1-15	4.86		1-14	4.43		1-20	5.07	
4 Year			4.89			5.99			5.59
Male	0-17	6.09		0-16	5.87		0-30	6.43	
Female	0-22	5.63		1-35	7.08		0-18	5.08	
5 Year			5.24			5.60			2.90
Male	1-23	7.53		0-20	7.21		0-9	4.32	
Female	0-8	4.00		1-15	5.08		0-8	3.00	
6 Year			5.56			4.67			3.29
Male	0-17	6.22		1-21	6.44		0-14	3.19	
Female	0-26	7.10		0-20	5.29		0-18	4.32	
7 Year			4.35			4.81			3.62
Male	1-15	6.89		1-15	6.39		0-21	3.75	
Female	0-21	5.76		0-20	6.69		0-13	4.03	
8 Year			7.07			5.13			4.25
Male	0-42	9.50		0-26	5.47		0-25	3.80	
Female	0-19	5.68		1-19	6.45		0-12	3.61	
Mean Standard Deviation			5.26			4.85			4.41

relatively stable. The mean standard deviations of the distributions of test to retest differences are shown at the bottom of Table 30. A single best over-all estimate of DIP test-retest reliability for clinical use is five items. That is, for the present the estimate of test reliability (the range within which two-thirds of the DIP test-retest differences would be expected to fall) is ± 5 items.

DIP Score Discrimination Curve. Preliminary inspection of the raw score data across sex groups when taking the DIP Forms 1, 2 and 3 did not indicate variability in shape of intelligibility curve as a function of presentation level. Therefore, it was not necessary to present a separate intelligibility curve for each sex sub-group.

Changes in DIP score as a function of presentation level for each of the age sub-groups indicated only slight deviations from linearity (although the curves were at higher score levels for the older age groups) and the curves were essentially similar in shape by virtue of having very similar slopes. The following slopes as a function of dB level were found for DIP Forms combined:

3 year group: 1.90 items per dB
4 year group: 1.80 items per dB
5 year group: 1.85 items per dB
6 year group: 1.83 items per dB
7 year group: 1.78 items per dB
8 year group: 1.75 items per dB

The overall mean of these rates is 1.8. Although a slight change in slope as a function of age may be suspected, the range of slope values is small. To require a separate slope for each age group would be impractical, and probably unnecessary.

For levels zero to ten decibels inclusive above speech reception threshold, the normal DIP test function may be taken as 1.8 items per decibel change in presentation level.

CHAPTER V

CONCLUSIONS AND DISCUSSION

The purpose of the present project was to prepare a set of two tests of the ability to hear speech (the Threshold by Identification of Pictures test, and the Discrimination by Identification of Pictures test) for clinical use with children. The preliminary forms of both tests were at hand prior to beginning this project. For the present study the test materials were further developed, test procedures were worked out, and data were collected and analyzed on 295 normal hearing children between the ages of three years and eight years inclusive.

The design of the study called for the testing of a number of hypotheses related to test form differences, sex differences, age factors, and test reliability. The following conclusions specify various aspects of the tests as do additional descriptive analyses of both of the tests, the test protocols, and the interpretation data presented in Appendix A. It is understood that the following is intended to apply only to normal children three to eight years old inclusive.

Conclusions Regarding TIP Test. The following conclusions were drawn regarding the Threshold by Identification of Pictures test:

1. Because the data did not indicate significant differences, in a number of instances null hypotheses related to the following were retained, and for the present it will be assumed there are not differences in:

a. TIP Test Forms A and B variability of threshold measurement among normal hearing children ages three to eight years.

b. TIP threshold levels between the sexes.

c. TIP threshold variability between the sexes, or among age groups.

2. TIP test Form B produces more intense (larger) threshold measurements than TIP test Form A for all ages and for both sexes, with the order of difference being between one and two decibels.

3. There is an age effect for both males and females, and for both TIP Forms A and B. TIP test threshold measurements show decreasing intensity levels with increasing age, to the extent of about eight decibels.

4. TIP test-retest threshold reliability, as indicated by product moment r, is positive and different from zero.

5. The TIP test threshold reliability expressed as decibel difference between test and retest thresholds for all age and sex subgroups, and for both TIP Forms is approximately three decibels. That is, the best estimate for test reliability (standard error of estimate) or range within which two-thirds of the test-retest scores lie is plus or minus three decibels.

6. The TIP test intelligibility curve (change in per cent items correct per decibel change in presentation level) is essentially equivalent for both Forms A and B, for both sexes, and for all age groups, three to eight years inclusive. The form of the intelligibility curve is specified by the following:

change from twelve per cent to 75 per cent intelligibility over a range of 14.5 dB (4.3 per cent intelligibility change per dB intensity change)

zero per cent intelligibility at twelve dB below decibel level which gives twelve per cent intelligibility

ninety per cent intelligibility at five dB above decibel level which gives 75 per cent intelligibility

one hundred per cent intelligibility at fifteen dB above decibel level which gives 75 per cent intelligibility

Conclusions Regarding DIP Test. The following conclusions were drawn regarding the DIP test:

1. Because the data did not indicate significant difference in a number of instances null hypotheses related to the following were retained, and for the present, it will be assumed there are not differences in:

a. DIP Forms 1, 2 and 3 in score obtained among children three to eight years of age, at any of the presentation levels SRT + 0 to SRT + 10 dB.

b. Variability of DIP scores for Forms 1, 2 and 3 among age and sex groups of children.

c. Level of DIP scores as a function of sex.

2. There is an increase in DIP scores with increased age both for males and females over the age range three to eight years.

3. DIP test-retest score reliability as indicated by product moment r is positive and different from zero.

4. The standard deviation of the distribution of absolute difference between test and retest scores on DIP Forms 1,2 and 3 is approximately five items. This value is the best estimate of

test-retest reliability for a subject, is the range within which two-thirds of test-retest scores may be expected to lie, and constitutes the standard error of measurement for the test.

5. The test score plotted against dB presentation level for the DIP Forms 1, 2 and 3, for both sexes and for age groups three to eight years inclusive, is 1.8 items change in test score per dB change in presentation level over the range SRT + 0 dB through SRT + 10 dB.

Comments Regarding Experimental Findings. Although a series of null hypotheses were stated in the original design for this study, it was expected that there would be a number of differences among sub-groups. It was anticipated that Forms A and B of the TIP test would be equivalent; the present findings indicate that more intense thresholds are obtained using Form B. However, this difference, being on the order of one and a half dB, either can be ignored for clinical test purposes inasmuch as it lies well within the usual expected test variability of approximately two and a half to three decibels for speech reception testing, or the clinician may use an appropriate correction factor between these two test Forms. (See Appendix A)

It was anticipated that female subjects would demonstrate more acute threshold measurements than male subjects. However, this was not borne out by the experimental data. Whether this indication is true of the population, that in auditory threshold females are not superior to males at these younger ages remains to be seen. The finding of no difference between the sexes either for test score or for variability in test scores was also found with the DIP test.

Probably the most significant positive experimental finding of this study, both for the TIP and the DIP tests is the significant age effects. That is, test performance improved (more acute thresholds and higher DIP scores) with increased age over the range three years to eight years. The change in threshold was about eight decibels; the change in DIP score was an increase of about nine in 48 items. Although the audiological literature, especially for pure tone testing, suggests that threshold changes as a function of age in young children, to our knowledge this has not heretofore been demonstrated for speech reception thresholds or for discrimination tests. This shift should be taken into account when calibrating a speech audiometer and when interpreting hearing test scores on children of various ages.

Although there was an increase in test performance with age, the shape of the growth curves obtained in the present study were somewhat surprising. In Appendix A is shown a diagram THRESHOLDS FOR AGE GROUPS ON TIP TEST. This was derived by using the original test data on subjects as a function of age, pooled for the sexes, but kept separate for test forms. Usually growth curves for such functions begin at a relatively poor level (high threshold) for the younger ages and approach a minimum as an asymptote curve for the older ages. However, our best estimate of the growth curves for the present data are relatively straight line threshold change from about the ages three through six, with some increasing rate of threshold change towards more acute thresholds in the age range six to eight. These curves argued against simple product moment correlation computations as a method of deriving curves of best fit. Although there was lack of linearity, the departure from linearity, seen as an increasing change in threshold rate for the older age levels, is no more than one or two decibels. Nevertheless, our present recommendation is that for the TIP test, the chart as appearing in the Appendix, and as part of the test book, be utilized for indicating best estimate of change in normal threshold as a function of age.

For experimental convenience the test item presentation for this research was recorded. Although the argument has been advanced that only by using recordings of speech test materials can adequately accurate work be done, available information comparing recorded and live voice presentation, using skillful announcers and suitable speech audiometers, indicates that test-retest reliability is as good using live voice as using recordings (Siegenthaler and Smith, 1961; Creston, J., Gillespie, M., and Krohn, C., 1966). A more important difference between live voice and recorded testing is the difference in obtained threshold. Siegenthaler and Smith (1961) tested 32 normal adults with PAL test No. 9 and with CID test W-2 using both live voice and recordings (of the same voice), and using both field testing and earphones monaurally and binaurally. The difference between thresholds obtained by recorded and by live voice testing ranged between 1.1 dB and 2.6 dB, the over-all best estimate of the difference being two decibels (live voice giving more acute threshold measurements). Our recommendations for the present TIP test materials is that if live voice rather than recorded test presentation be used, two decibels be subtracted from the thresholds indicated in Appendix A chart THRESHOLDS FOR AGE GROUPS ON TIP TEST.

In the case of the DIP test scores at all three presentation levels, the growth curves were more conventional than for the TIP test in that they showed increased scores as a function of age, with indications of leveling off at the upper age levels. In Appendix A the chart SCORES FOR AGE GROUPS ON DIP TEST presents our present best estimate of normal DIP scores as a function of age for the sexes combined at each of the three presentational levels used. No data are available on live voice versus recorded presentation of the DIP test.

When using both of the normative charts in Appendix A it is recommended that children at between yearly age levels be given threshold estimates by interpolation.

The smooth curves for both of the charts in Appendix A were obtained by visual curve fitting for least sum of deviations, using the intermediate step of connecting successive midpoints as the first curve-smoothing technique. The curves are not intended to be projected below the age of three nor above the age of eight.

The major product of clinical importance from this study is the accompanying material which presents the TIP and DIP test pictures and test protocols (Siegenthaler and Haspiel, 1966). Because it was judged impractical to bind the pictures together with the present research report, they were excluded. However, the test protocols, together with related data for clinical use are included as Appendix A to this research report.

Our experience to date with these materials, and the experience of a number of other audiology clinics which have used these tests (the test pictures being produced either by hand drawing individually or by cutting pictures and pasting them on appropriate cards) indicates the usefulness of the procedures for young children in the audiology clinic. This includes not only the ability to utilize the pictures successfully for obtaining of what appear to be reliable and satisfactory estimates of hearing ability in children, but also includes the strong clinical impressions that the obtained scores are meaningfully related to the child's audiological status.

CHAPTER VI

SUMMARY

A continuing problem in clinical audiology is the hearing testing of young children, especially with respect to their ability to hear speech. Existing speech reception threshold test materials for children usually are only versions of materials suitable for adults. One of the present authors assisted in earlier work which developed the preliminary form of a speech reception test for children, utilizing a picture indentification procedure. This test procedure demonstrated its usefulness with children, but it was not available in readily useful form, nor were satisfactory standardization data available. A part of the present project was to develop further this speech reception test, here called the Threshold by Identification of Pictures test (TIP).

There is an even greater absence of suitable speech reception tests of discrimination for children. Earlier work by the present authors laid the groundwork for, and made a preliminary draft of, a Discrimination by Identification of Pictures test (DIP) suitable for children. This test was not based upon phonetic balance, which has not been verified as important to a speech hearing discrimination test. Rather, it utilized the phonetic elements of voicing of consonants, pressure pattern of consonants, and transitional patterns between adjacent phonemes. These factors have been demonstrated to be related to speech intelligibility by a number of investigators.

Both of the above tests only require that the child point to pictures suitable for his age, upon hearing the names of the pictures. The TIP test has a total of 25 test items in each of its two forms; the DIP test has a total of 48 items for each of its three forms. Both tests are administered over the normal speech audiometer.

For the present project 295 normal children ages three to eight years of age were tested. There was an approximate equal distribution between the sexes, and among yearly age groups. All subjects passed an otological inspection, pure tone threshold audiogram, and had intelligence quotients between 90 and 110. Subjects were tested on both forms of the TIP test for threshold and at three intensity levels for discrimination on the DIP test. Each subject was given a complete re-test within one week to obtain reliability data.

Data analysis was for test-retest reliability, sex differences, age norms, test form differences, and for shape of intelligibility curves.

The following conclusions were drawn regarding the Threshold by Identification of Pictures test:

1. There are not differences in TIP Forms A and B in variability of threshold, in TIP thresholds between the sexes, or in TIP threshold variability between the sexes or among age groups.

2. TIP test Form B produces more intense threshold measurement values than TIP test Form A for all ages and for both sexes, with the order of difference being between one and two decibels.

3. There is an age effect for both males and females, and for both TIP Forms A and B. TIP test threshold

measurements show decreasing intensity levels with increasing age, to the extent of about eight decibels over the age range three to eight years.

4. The TIP test threshold reliability, expressed as decibel difference between test and retest, is about three decibels. That is, the best estimate of test reliability, or range within which two-thirds of the test-retest scores lie, is plus or minus three decibels.

5. The TIP test intelligibility curve is essentially equivalent for both Forms A and B, for both sexes, and for all age groups three to eight years inclusive. The form of the curve is specified as:

   change from twelve to 75 per cent intellibility over a range of 14.5 dB
   zero per cent intelligibility at twelve dB below level which gives twelve per cent intelligibility
   ninety per cent intelligibility at five dB above level which gives 75 per cent intelligibility
   one hundred per cent intelligibility at fifteen dB above dB level which gives 75 per cent intelligibility.

The following conclusions were drawn regarding the Discrimination by Identification of Pictures test:

1. There are not differences in DIP score for Forms 1, 2 or 3 among children three to eight years of age at any of the presentation levels SRT + 0, +5, or +10 dB, nor in variability of DIP scores among test forms for age or sex groups of children, nor in level of DIP test score as a function of sex.

2. There is an increase of DIP score with increased age, over the range three to eight years.

3. The standard deviation of the distribution of absolute difference between test and retest scores on DIP Forms 1, 2, and 3 is approximately five items. This value is the best estimate of test-retest reliability, is the range within which two-thirds of test-retest scores may be expected to lie, and constitutes the standard error of measurement for the test.

4. The test score plotted against presentation level of the DIP Forms 1, 2 and 3, for both sexes and for age groups three to eight years inclusive, is 1.8 items change in test score per dB change in presentation level over the range SRT + 0 dB through SRT + 10 dB.

Appended to the full research report, and a product of the project, is the test protocol and test interpretation data for TIP and DIP. Separately produced and bound, together with the test protocols and interpreation data are the test pictures. Our experience to date with these materials, and the experience of a number of other audiology clinics which have used these tests on a trial basis indicates the usefulness of the procedures for young children. This includes not only the ability to utilize the pictures successfully for obtaining what appear to be reliable and satisfactory estimates of hearing ability in children, but also includes the strong clinical impressions that the obtained scores are meaningfully related to the child's audiological status.

REFERENCES

Black, J., "Accompaniements of Word Intelligibility", Journal Speech and Hearing Disorders, 17, 409-418 (1952).

Carnhart, R., "Speech Audiometry", Acta-Otolaryngologica, 1-42 (1953).

Carnhart, R., Tillman, T., and Wilber, L., "A Test for Speech Discrimination Composed of CNC Monosyllabic Words", Perceptual and Motor Skills, 16, 680 (1963).

Cooper, F. S., Delattre, P. D., Liberman, A. M., Borst, J. M., and Gerstman, L. J., "Some Experiments on the Perception of Synthetic Speech Sounds", Journal Acoustical Society of America, 24, 597-606, (1952).

Creston, J., Gillespie, M., and Krohn, C., Speech Audiometry: Taped vs. Live Voice, AMA Archives of Otolaryngology, 83, 14-17 (1966).

Delattre, P., Liberman, A., Cooper, F., "Acoustic Loci and Transitional Cues for Consonants", Journal Acoustical Society of America, 24, 769-773 (1955).

Egan, J., "Articulation Testing Methods", OSRD Report 3802, United States Department of Commerce, Office of Technical Services, Washington, D.C., (1944).

Ewing, I.R., "The Ascertainment of Deafness in Infancy and Early Childhood", Journal Laryngology and Otology, 59, 137-141 (1943).

Fletcher, H., and Steinberg, J., "Articulation Testing Methods", Bell System Technical Journal, 8, 806-854 (1929).

Fowler, E. Sr., "Hearing Standards for Acceptance Disability Rating and Discharge in the Military Services and in Industry", Laryngoscope, 57, 937-956 (1941).

Goldstein, L., Intelligibility Scores Obtained with Shortened Phonetically Balanced Word Lists, M.S. Thesis, The Pennsylvania State University, (1965).

Hanley, C.N., "Factorial Analysis of Speech Perception", Journal Speech and Hearing Disorders, 21, 76-87 (1956).

Hansen, B.F., "The Application of Sound Discrimination Tests to Functional Defectives", Journal Speech and Hearing Disorders, 9, 347-55 (1944).

Harris, J.H., "Some Suggestions for Speech Reception Testing", *AMA Archives of Otolaryngology*, 50, 388-405 (1949).

Haskins, H. L., *A Phonetically Balanced Test of Speech Discrimination for Children*, M.A. Thesis, Northwestern University (1949).

Haspiel, G.S., *Picture Identification Test of Discrimination for Speech*, Ph.D. Thesis, The Pennsylvania State University (1961).

Hirsh, I.J., *The Measurement of Hearing*, McGraw-Hill Co., New York (1952).

Hirsh, I., Davis, H., Silverman, S., Reynolds, E., Eldert, E., Benson, R., "Development of Materials for Speech Audiometry", *Journal Speech and Hearing Disorders*, 17, 321-337 (1952).

Howes, D., "On the Relation Between the Intelligibility and Frequency of Occurrence of English Words", *Journal Acoustical Society of America*, 29, 296-305 (1957).

Hudgins, C., "A Method of Appraising the Speech of the Deaf, *Volta Review*, 5, 597-601 (1949).

Joos, M., "Acoustic Phonetics", *Language*, Monograph Supplement No. 23, (1948).

Karlin, J.E., "A Factorial Study of Audience Function", *Psychometrika*, 7, 251-279 (1942).

Keaster, J., "A Quantitative Method of Testing the Hearing of Young Children", *Journal Speech and Hearing Disorders*, 12, 159-162 (1947).

Liberman, A., Delattre, P., Gerstman, L. and Cooper F., "Tempo of Frequency Change as a Clue for Distinguishing Classes of Speech Sounds", *Journal Experimental Psychology*, 52, 127-137 (1956).

Liberman, A., Ingeman, F., Lisker, L., Delattre, P., and Cooper, F., "Minimal Rules for Synthesizing Speech", *Journal Acoustical Society of America*, 31, 1490-1499 (1959).

Lindquist, E., *Design and Analysis of Experiments in Psychology and Education*, Houghton Mifflin Co., Boston, pp. 281-284, 292-297, (1953).

Mange, C., *A Study of Speech Sound Discrimination with Words and Sentences*, M.S. Thesis, The Pennsylvania State University (1953).

Mansur, R. W., The Construction of a Picture Test for Speech Sound Discrimination, M. Ed. Thesis, Boston University (1950).

Mason, H., "Phonetic Characteristics Related to Intelligibility of Words in Noise", Psychological Corps Bulletin, OSRD 5174, 6 (1945).

Meyerson, L., "A Verbal Audiometric Test for Young Children", American Psychologist, 2, 291 (1947).

Miller, G., Heise, G., and Lichter, W., The Intelligibility of Speech as a Function of Test Material, Journal Experimental Psychology 48, 329-335 (1951).

Monsees, E.K., "Children's Auditory Test", Volta Review, 446 (1953).

Morley, D., "Rationalism in Testing the Hearing of Children", Volta Review, 468-478 (1948).

Mullen, W., Relationships among Tests of Pure Tone Threshold, Speech Hearing, Auditory Discrimination, and Articulation, M.S. Thesis, Pennsylvania State University (1954).

Nasca, F.L., An Investigation of the PICSI Test: Picture Indentifica-Test for Children of Standard Index, Ph. D. Thesis, Indiana University (1964).

Owens, E., "Intelligibility of Words Varying in Familiarity", Journal Speech and Hearing Research, 4, 113-129 (1961).

Oyer, H. and Doudna, M., "Word Familiarity as a Factor in Testing Discrimination", AMA Archives of Otolaryngology, 72, 351-355 (1960).

Pearson, J., A Speech Reception Test for Young Children, M.A. Thesis, University of Michigan (1951).

Peters, R., "Dimensions of Perception for Consonants", Journal Acoustical Society of America, 35, 1963, 1985-1989 (1963).

Potter, R., Kopp, G., and Green, H., Visible Speech, D. Van Nostrand Company, Inc. (1947).

Pronovost, W. and Dumbleton, C., "A Picture-Type Speech Sound Discrimination Test", Journal Speech and Hearing Disorders, 18, 258-266 (1953).

Resnick, D., "Reliability of the Twenty-five Word Phonetically Balanced Lists", Journal Auditory Research, 2, 5-12 (1962).

Rosenzweig, M., and Portman, L., "Intelligibility as a Function of Frequency of Usage", Journal Experimental Psychology, 54, 412-422 (1957).

Schiefelbush, R. and Lindsey, M.J., "A New Test of Sound Discrimination", Journal Speech and Hearing Disorders, 23, 153-159 (1958).

Siegenthaler, B., "Perception of Certain Phonetic Variables", Speech Monograph, 21, 39-45 (1954).

Siegenthaler, B. and Haspiel, G., TIP and DIP Test Materials: A Manual of Procedures and Test Pictures, U.S Office of Ed. Cooperative Research Program, Project No. 2372 (1966).

Siegenthaler, B., Pearson, J., and Lezak, R., "A Speech Reception Threshold Test for Children", Journal Speech and Hearing Disorders, 19, 360-366 (1954).

Siegenthaler, B. and Smith, D., "Speech Reception Thresholds by Different Methods of Test Administration", Journal Acoustical Society of America, 33, 1802 (1961).

Siegenthaler, B. and Strand, R., "Audiogram Average Methods and SRT scores", Journal Acoustical Society of America, 36, 583-589 (1964).

Silverman, S. and Walsh, T., "Diagnosis and Evaluation of Fenestration", Laryngoscope, 56, 536-555 (1946).

Sims, A., "Word List Picture Identification Test for Speech-Hearing in Preschool Children", M.S. Thesis, Louisiana State University (1961).

Solomon, L.N., Webster, J.C., and Curtis, J.F., "A Factorial Study of Speech Perception", Journal of Speech and Hearing Research, 3, 101-107 (1960).

Sortini, A. and Flake, C., "Speech Audiometry for Pre-School Children", Laryngoscope, 63, 991-997 (1953).

Sparks J., "Expository Note on the Problem of Making Multiple Comparisons in a Completely Randomized Design, Journal Experimental Education, 31,343-349 (1963).

Spilka, B., "The Relationship between Talker Intelligibility and Message Familiarity", SOC Test Rehabilitation 104-2-30 (1954).

Templin, M., "Speech Development in the Young Child: 3. The Development of Certain Language Skills in Children", Journal Speech and Hearing Disorders, 17, 280-285 (1952).

Thwing, E., "Effects of Repetition on Articulation Scores", Journal Acoustical Society of America, 28, 302-303 (1956).

Tobias, J., "On Phonemic Analysis of Speech Discrimination Tests", Journal Speech and Hearing Research, 7, 13-22 (1965).

Travis, L. and Rasmus, B., "The Speech Sound Discrimination Ability of Cases with Functional Disorders of Articulation", Quarterly Journal of Speech, 17, 217-226 (1931).

APPENDIX A

PROTOCOL FOR TIP AND DIP TESTS

(Taken from *TIP and DIP Test Materials: A Manual of Procedures and Test Pictures* which includes the test pictures.)

TIP and DIP Test Materials:

A Manual of Procedures and Test Pictures

Bruce M. Siegenthaler, Ph.D.

and

George S. Haspiel, Ph.D.

Speech and Hearing Clinic
Department of Special Education
The Pennsylvania State University
University Park, Pennsylvania
30 June, 1966

These materials were developed as part of a research project supported by the Cooperative Research Program of the Office of Education, U.S. Department of Health, Education and Welfare, project No. 2372, contract No. OE 5-10-003. The research report is Development of Two Standardized Measures of Hearing for Speech by Children (1966).

TABLE OF CONTENTS

Page

INTRODUCTION.................... III

THRESHOLD BY IDENTIFICATION OF PICTURES (TIP) TEST.................... IV

- Description of TIP Test.................... IV
- Instructions for Administering TIP Test.................... V
- TIP Scoring, Tabular Method.................... VI
- TIP Scoring, Graphic Method.................... VI
- Interpretation of Scores.................... VII
- TIP SCORING SHEET.................... X
- THRESHOLDS FOR AGE GROUPS ON TIP TEST.................... XI
- TIP TEST CARDS....................
  - TIP A Practice
  - TIP A...... 1-5
  - TIP B Practice
  - TIP B...... 1-5

DISCRIMINATION BY IDENTIFICATION OF PICTURES (DIP) TEST.................... XII

- Description of DIP Test.................... XII
- Instructions for Administering DIP Test.................... XIII
- DIP Scoring.................... XIV
- Interpretation of Scores.................... XIV
- DIP Sample Scoring Sheet.................... XVI
- DIP Test Scoring Table.................... XVII
- Scores for Age Groups on DIP Test.................... XVIII
- DIP TEST CARDS....................DIP A, B, C, D, Practice
  - DIP V............... 1-5
  - DIP I............... 6-9
  - DIP P............... 10-17
  - DIP VI............... 18-25
  - DIP VP............... 26-29
  - DIP IP............... 30-37
  - DIP VIP............... 38-48

THRESHOLD BY IDENTIFICATION OF PICTURES (TIP) TEST
AND
DISCRIMINATION BY IDENTIFICATION OF PICTURES (DIP) TEST

Introduction

The TIP and DIP tests are measures of speech hearing ability constructed and validated for use with pre-school and school age children. The tests may be administered in the usual audiologic facility, using equipment and procedures common to threshold and discrimination tests employing verbal materials.

The child should sit at a previously determined point relative to the sound source at a small table appropriate for his size. When using live voice procedures an assistant to the tester may sit beside the child, turning the test picture pages, and if necessary indicating to the tester when the next item may be presented. Either the assistant or tester may mark through the incorrectly identified item on the scoring sheet for a permanent record, or keep a tally using a hand counter.

These tests may be administered using live voice, or recorded materials can be prepared. Tests may be administered in a sound field, or through earphones, either monaural or binaural.

The original research was done using recorded materials in a sound field. The accompanying norms are for this condition. Live voice presentation has the advantage of greater flexibility for meeting the child's needs and has no poorer reliability than recorded presentation, using skillful testers and a speech audiometer. Live voice produces two dB more acute threshold measurements than recorded. Binaural earphones (if the ears are equal) give three dB more acute threshold measurements than monaural earphone. Field testing (binaural) gives three dB more measured acuity than binaural earphones. These values should be taken into account when interpreting a TIP test threshold. No similar values are available for intelligibility testing or for discrimination testing as with the DIP test.

Although only very familiar words have been included, it is possible that some hearing impaired children will be unfamiliar with some of the items. Before beginning either of the tests it is advisable to determine if there are unfamiliar items; if so, they should be explained and the names taught to the child. Usually the parent can indicate items unfamiliar to the child.

Threshold by Identification of Pictures (TIP) Test

Description of TIP Test

The TIP test consists of two Forms (Form A and Form B) of multiple choice items for obtaining speech reception threshold. Each form has a practice card and five test cards. There are five items on each card. The items are monosyllabic familiar word-pictures which are called to the child in a predetermined order following the usual speech reception test procedures. A child responds by pointing to the picture of the word he thought he heard.

Instructions for Administering TIP Test

1. Check child's familiarity with test items.
2. Explain test procedure face-to-face, or use speech audiometer.
3. Use TIP Scoring Sheet for lists of call words for each TIP picture card.
4. Explore for approximate speech reception threshold using practice card TIP A.
    a. Monitor carrier phrase "Point to the . . ." and the test items at VU O.
    b. Call first Practice item at a level well above expected threshold.
    c. Continue calling Practice items, attenuating 10 dB after each item until child consistently fails to respond correctly.
    d. Re-use Practice items on the same card as necessary.
    e. Add 10 dB to last correct-response level on Practice items, and write this beginning level value on the scoring sheet in the top space under audiometer setting. Write the presentation levels for the other four words in each column in 5 dB decreasing steps in the spaces provided.
5. Present first item on card TIP A-1 at the beginning level as noted on the top line under audiometer setting, and observe child's pointing response.
6. Call second item in the column at a level 5 dB less than the first item (as previously noted under audiometer setting). Observe response, and call third item 5 dB lower than the preceding item. Continue in 5 dB attenuation steps until all five words for a card are called. Draw a line through items not identified (pointed to) correctly. Call each item only once; even though the child does not respond correctly go directly to the next item at a 5 dB lower level. (Only exception is gross extenuating circumstance).
7. Turn to card TIP A-2 and give the first item at the same level as the first item on card TIP A-1. Use same procedure as before for remaining items on card TIP-2 (i.e., attenuating 5 dB after each item).
8. Present cards A-3, A-4 and A-5 to complete TIP Form A. Give the first item on each card at the beginning level, and attenuating 5 dB per item for a given picture card.

9. It is necessary that the first item of every card be identified correctly and the last item be failed if the tabular method of scoring is to be employed. In some cases, however, it will not be possible to achieve this and the graphic method may be used. If the beginning level was not estimated correctly for the first card, the beginning level may be adjusted for subsequent cards. Retest on first card after the other four cards have been used.

10. Tell the child to relax until it is time for the next test.

11. TIP Form B is administered the same as TIP Form A, use the Form B cards and call words.

12. Individual cards from one TIP Form are not interchangeable with individual cards from the other Form.

TIP Scoring, Tabular Method (done for each form separately)

1. Count number of items correctly identified across all cards for a TIP Form at each audiometer setting, and write number in space to the right on the Scoring Sheet under NUMBER CORRECT for each level separately. Total the number of items identified correctly at all settings on all cards, and write the number in the space TOTAL CORRECT. (A tally counter may be used if the tester is certain the beginning level will produce all correct responses.

2. Subtract TOTAL CORRECT from speech audiometer attenuator dB setting for beginning level of audiometer setting.

3. Add 2.5 dB correction factor to obtain speech reception threshold (to compensate for the 5 dB attenuation steps used.)

4. The THRESHOLD may be recorded on the scoring sheet in the space provided.

5. Thresholds may be noted in dB re .0002 dy/cm^2, or as hearing levels. (See below.)

TIP Scoring, Graphic Method (done for each form separately)

1. Count the number of items correctly identified across all cards for a TIP Form at each

presentation level, and write number in space to the right on the Scoring Sheet under NUMBER CORRECT for each presentation level separately.

2. Write dB presentation levels at bottom of SCORING GRID on Scoring Sheet for a series of five levels putting lowest dB level to the left and using one vertical line per 5 dB step. (The grid provides space for repeated test presentation.)

3. For each dB presentation level mark number correct on the GRID.

4. Draw a smooth curve of best fit, using visual estimates (projecting the curve to five correct and to none correct) according to expected articulation curve shown on the TIP Scoring Sheet GRID and specified as:

   change from 12% to 75% correct over a range of 14.5 dB
   0% correct at 12 dB below dB level for 12% correct
   90% correct at five dB above dB level for 75% correct
   100% correct at 15 dB above dB level for 75% correct

   (This curve may be made into template form for ease of drawing.)

5. Observe intersection of 50% correct and the curve of best fit.

6. Project this intersection downward to the dB presentation level scale and note dB level (estimate to the nearest one-half decibel). This is the speech reception threshold.

7. The THRESHOLD may be recorded on the scoring sheet in the space provided.

8. Thresholds may be noted in dB re .0002 dy/cm^2, or as hearing levels. (See below.)

Interpretation of Scores

Refer to the chart THRESHOLDS FOR AGE GROUPS ON TIP TEST, which shows normal TIP thresholds using recorded materials in a sound field re .0002 dy/cm^2. If your speech audiometer is calibrated in dB re .0002 dy/cm^2 and you used recorded materials in a sound field, compute the child's speech reception threshold deviation from normal (his hearing level) by subtracting the appropriate dB correction factor (obtained on the left of the chart) for his age and for TIP test Form from the audiometer panel setting.

If your speech audiometer has other than an SPL calibration or if you use earphones or live voice testing, do not expect that your dB panel settings will produce threshold for the TIP test as shown on the chart THRESHOLDS FOR AGE GROUPS ON TIP TEST because of the effects of age and of test procedure. Calibrate your audiometer. When calibrating your speech audiometer the obtained normal SRT will be influenced by the test situation, in addition to the age of the children and TIP Form used. Live voice testing gives measured thresholds two dB lower in intensity than recorded voice; two earphone listening (assuming equal ears) gives three dB better measured acuity than one earphone listening; field testing gives three dB better measured acuity than binaural earphone listening.

Do a real-ear calibration, using at least ten children with known normal hearing, and all within three months of the same age. Obtain the average (median) speech audiometer panel setting in dB for their thresholds, and write this value to the right of the chart THRESHOLDS FOR AGE GROUPS ON TIP TEST for the test Form used and at the level of intersection of their age group line and the normal curve for the TIP test Form used. Interpelate dB levels when necessary. The appropriate one dB steps should be written along the right margin of the chart to complete the calibration corrections (larger dB values toward the top of the chart and smaller dB values toward the bottom).

As subsequent children are tested, obtain SRT and enter the chart at the bottom for the child's age, read up to the test Form used, and read right to find the appropriate correction factor. Subtract this value from the audiometer setting for threshold to obtain his hearing level (threshold deviation from normal) for his age and for test Form used.

Standard error of estimate for TIP thresholds is ± 3 dB. That is, 2/3 of the retest thresholds will be expected to fall within three decibels of initial test thresholds.

THRESHOLD BY IDENTIFICATION OF PICTURES (TIP) SCORING SHEET

NAME ____________________ SEX ______ DATE __________ EXAMINER __________

ADDRESS ______________________________ BIRTHDATE ______________

Field _____ Earphones _____ Test Validity _____ Remarks ____________________

FORM A

Practice A	audiometer setting	A-1	A-2	A-3	A-4	A-5	NUMBER CORRECT
chair	________	hands	house	horse	clock	eyes	______
shoe	________	dog	flag	ball	mouse	train	______
chair	________	truck	dress	cat	drum	milk	______
bread	________	stove	bus	fish	cup	eyes	______
cap	________	hands	kite	ball	clock	spoon	______

TOTAL CORRECT ______

FORM B

Practice B	audiometer setting	B-1	B-2	B-3	B-4	B-5	NUMBER CORRECT
shoe	________	door	bed	lamb	skates	tree	______
chair	________	pie	boat	corn	clown	knife	______
car	________	cow	hand	top	comb	doll	______
swing	________	socks	gum	lamb	gun	tree	______
dog	________	pie	bed	frog	clown	blocks	______

TOTAL CORRECT ______

THRESHOLD

Right_____dB

Left _____dB

Bin. _____dB

SCORING GRID

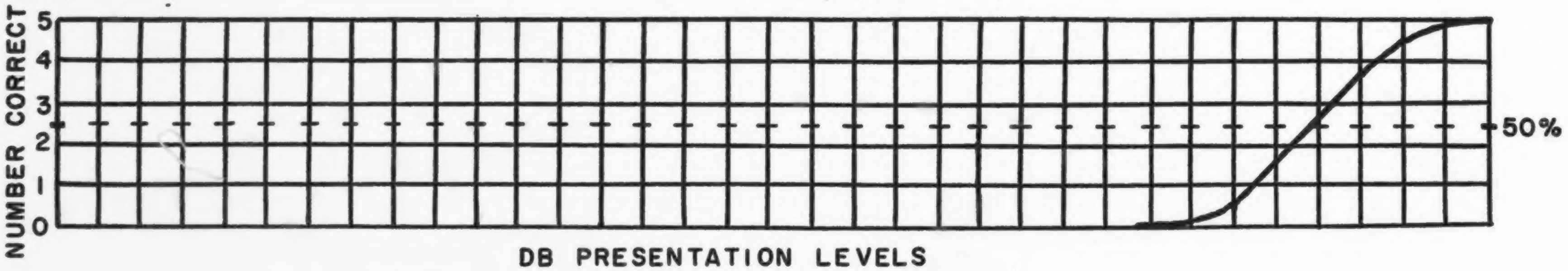
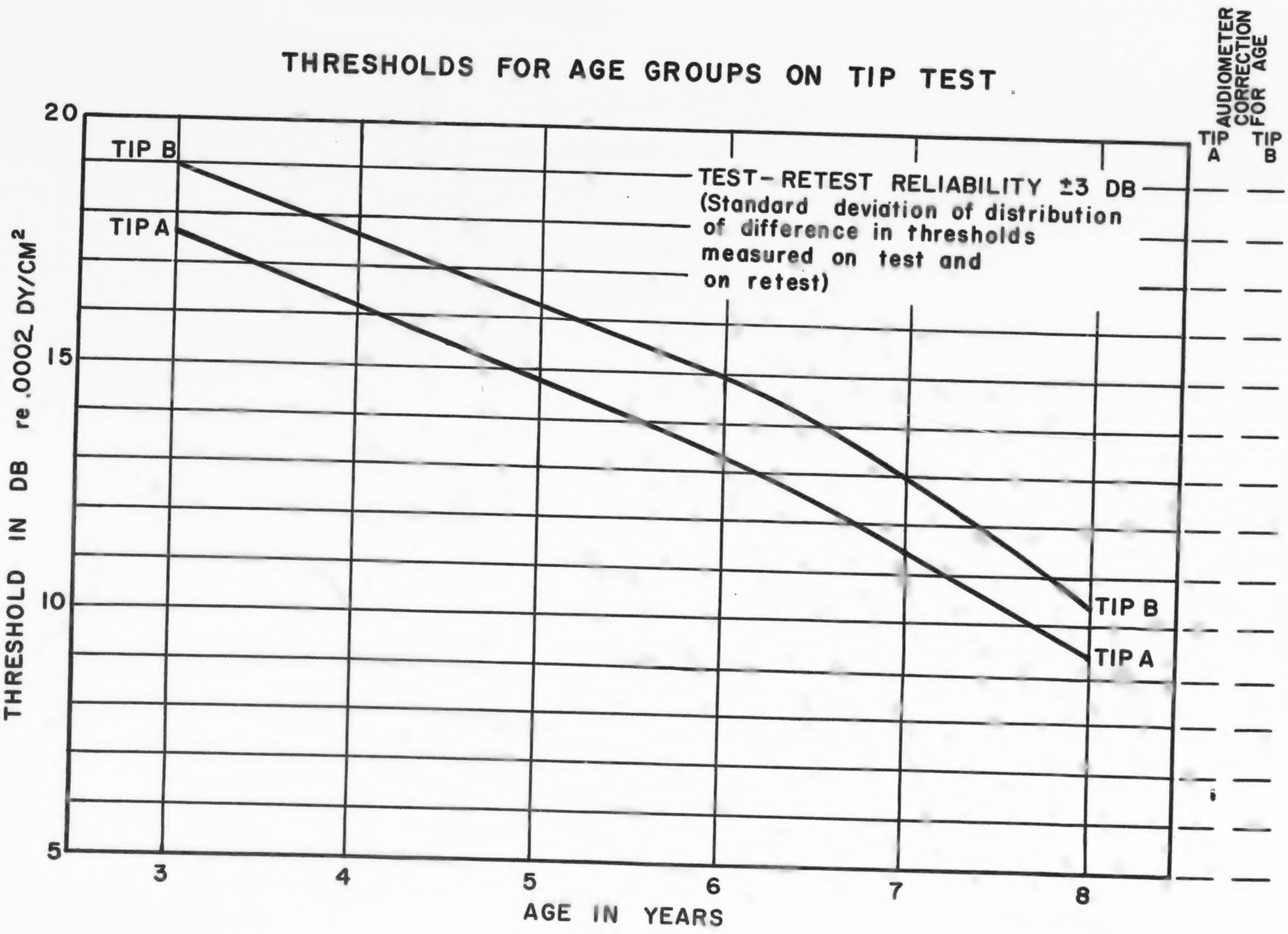


Figure 1. TIP test protocol and scoring s X

THRESHOLDS FOR AGE GROUPS ON TIP TEST



Discrimination by Identification of Pictures (DIP) Test

Description of DIP Test

The DIP test is composed of 48 paired monosyllabic items of which the initial consonants are different in voicing, influence, pressure pattern or combinations of these acoustic factors. The words used are familiar, and together with the pictures were validated for familiarity and recognition with a group of hearing impaired pre-school children.

Three Forms of the DIP test have been prepared using the pool of 48 pairs of pictures and words. Each of the Forms contains the following distribution of paired items differing in their initial consonants:

5 voicing (DIP pages VI-5)
4 influence (DIP pages I 1-5)
8 pressure pattern (DIP pages P 10-18)
8 voicing and influence (DIP pages VI 18-25)
4 voicing and pressure pattern (DIP pages VP 26-29)
8 influence and pressure pattern (DIP pages IP 30-37)
11 voicing, influence, and pressure pattern (DIP pages VIP 38-48)

The DIP test cards are used in the same order for all three forms.

The DIP test can be used for several audiological purposes including the measuring of speech discrimination ability at most-comfortable listening level, conversational level, or other appropriate points on the intensity scale. Data have been determined for normal hearing young children at SRT +

0 dB, SRT + 5 dB, and at SRT + 10 dB (speech reception thresholds obtained with the TIP test).

Instructions for Administering DIP Test

1. Check child's familiarity with test materials.

2. Orient child to test procedure, either face-to-face or over speech audiometer.

3. Select appropriate speech audiometer output level, either field or earphones, monaural or binaural and DIP Form. (All three DIP Forms are equivalent, and each has the same manner of presentation.)

4. Monitor voice at VU O for carrier phrase "Point to the . . ." and for DIP words.

5. Call first practice item, from DIP Scoring Sheet, which contains Practice items, and wait for response.

6. Child indicates item by pointing.

7. Call additional practice items, reusing them if necessary until child is familiar with the test procedure, and turning page after each item to use the four practice items.

8. Call first DIP test item, observe response, turn page, and call next item (at same intensity level).

9. Draw line through test word on DIP Scoring Sheet if incorrectly identified. (An alternative procedure is to use a hand tally counter, but this does not give a permanent record of specific items missed.)

10. Proceed in like manner to call remaining test items, without changing presentation level to complete all 48 items on the Form.

11. Call each item only once. Do not repeat unless gross extenuating circumstances require it; then repeat the item immediately.

12. When the DIP list is completed, tell the child to relax until the next test.

DIP Scoring

1. Count the number of items missed to find Error Score, and record on DIP Scoring Sheet.
2. Multiply error score by 2. 1 to determine Error Per Cent.
3. Subtract Error Per Cent from 100% to obtain DIP Per Cent.

(According to this method, a perfect discrimination score would be 100. 8%. For most purposes this discrepancy can be ignored. The DIP per cent may be read from the accompanying DIP Scoring Table which gives the DIP per cent for error scores, rounding to the nearest whole per cent.)

If an intelligibility curve is to be drawn showing DIP test score as a function of presentation level, the slope of the function is 1. 8 test items per dB change in presentation level over the range SRT + 0 dB through SRT + 10 dB.

Interpretation of Scores

Refer to normative data on SCORES FOR AGE GROUPS ON DIP TEST chart for comparison of child's score to normal for his age group at the presentation level used. These normative values were obtained with recorded voice, field presentation, on normal hearing children. Data are not yet available for older or younger children or for other manners of test administration. The present best estimate is that there is little change in DIP scores after eight years of age.

Because 50% of the items in the DIP test could be guessed by chance, the range of critical scores is from 50% to 100%; a 50% or less score is a poor score. It is important that the scores be described as DIP Scores because of the possible confusion with other tests based on different principles and requiring different forms of response.

The standard error of estimate for the DIP Score is ± 5 items. That is, 2/3 of the retest scores will be expected to fall within five items (approximately 10%) of the test score.

Discrimination by Identification of Pictures (DIP)

Scoring Sheet

NAME ______________________________ AGE ________ DATE ____________

Card	DIP Form 1	DIP Form 2	DIP Form 3
Practice A	dog	cat	dog
B	chair	boat	chair
C	kite	kite	kite
D	coat	kite	coat
DIP V-1	bear	pear	pear
2	deer	tear	deer
3	bees	peas	peas
4	fan	man	man
5	coat	goat	coat
I-6	key	pea	key
7	pup	cup	cup
8	goat	boat	goat
9	pea	tea	tea
P-10	meat	beet	meat
11	saw	paw	saw
12	chain	cane	cane
13	wheel	seal	wheel
14	cheese	keys	keys
15	wing	ring	ring
16	rat	bat	rat
17	sail	tail	tail
VI-18	coat	boat	coat
19	bow	toe	toe
20	toy	boy	boy
21	back	tack	back
22	cat	bat	cat
23	pot	dot	dot
24	bone	cone	bone
25	bee	key	bee
VP-26	nail	sail	sail
27	men	pen	men
28	sun	gun	sun
29	feet	beet	feet
IP-30	cat	hat	hat
31	fire	tire	fire
32	horn	corn	horn
33	pear	hair	hair
34	dog	log	log
35	can	fan	can
36	peas	cheese	peas
37	shoe	two	shoe
VIP-38	light	kite	kite
39	cheese	bees	bees
40	rose	toes	rose
41	rain	cane	rain
42	hat	bat	bat
43	gum	thumb	gum
44	log	hog	log
45	four	door	four
46	man	can	can
47	suit	boot	suit
48	bear	hair	hair

100% minus Error Per Cent : ____ %

DIP Per Cent : ____ %

DIP Error Score : ____ x 2.1

Error Per Cent : ____ %

100% minus Error Per Cent : ____ %

DIP Per Cent : ____ %

DIP Error Score : ____ x 2.1

Error Per Cent : ____ %

100% minus Error Per Cent : ____ %

DIP Per Cent : ____ %

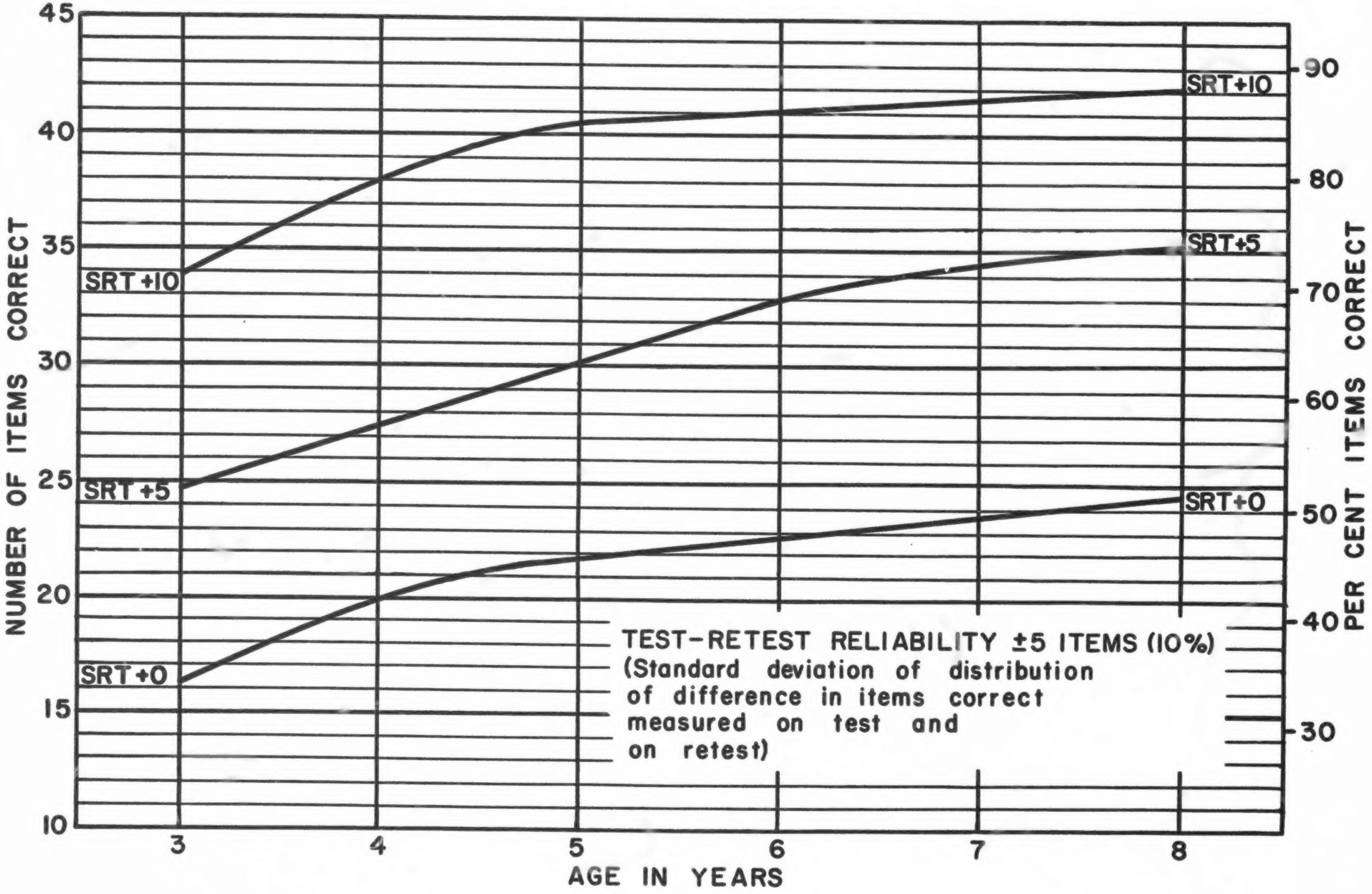
DIP Error Score : ____ x 2.1

Error Per Cent : ____ %

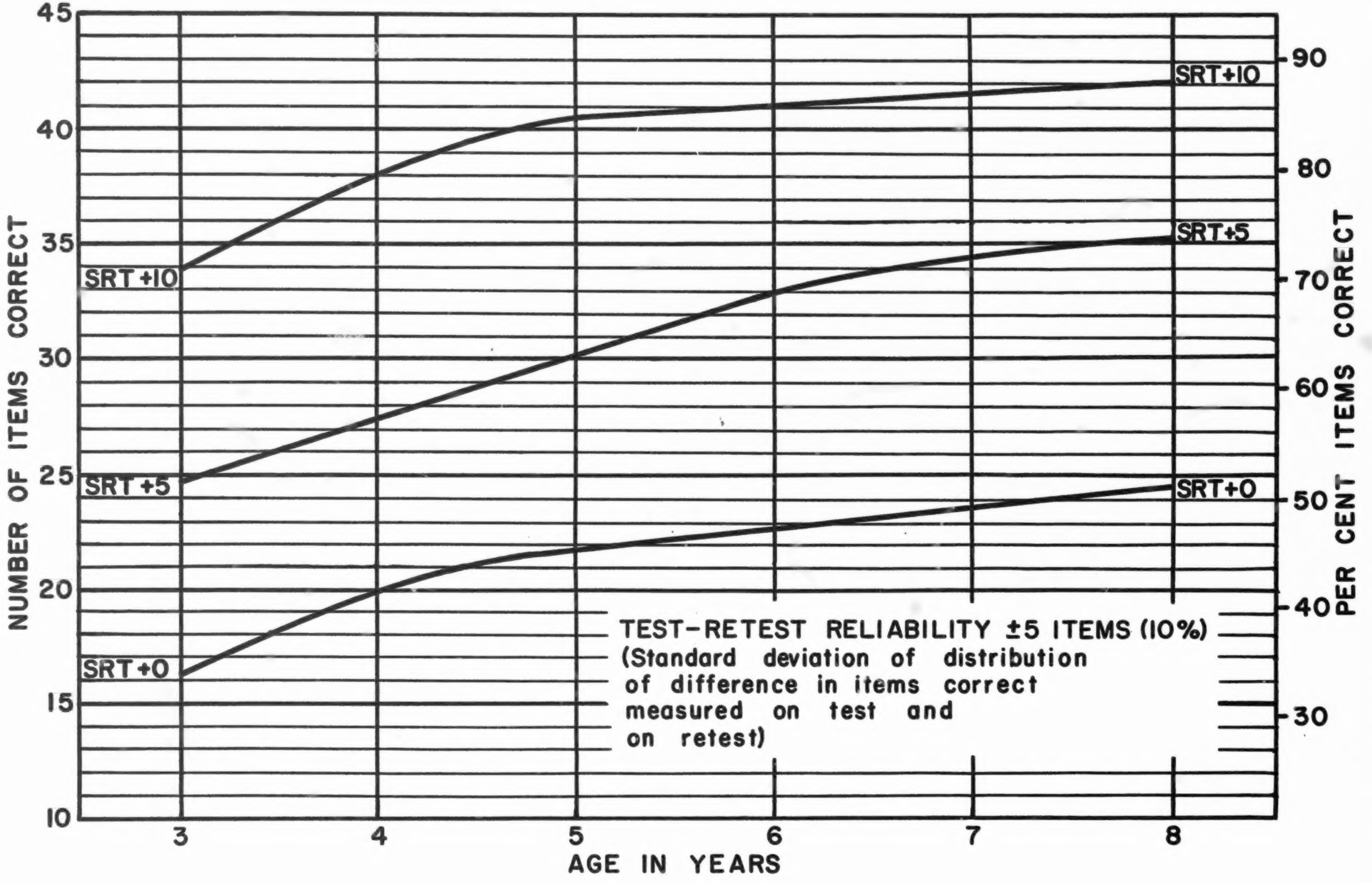
DIP test scoring table: Error Score converted to DIP per cent correct

Error Score	DIP Per Cent	Error Score	DIP Per Cent
0	100	19	60
1	98	20	58
2	96	21	56
3	94	22	54
4	92	23	52
5	89	24	50
6	87	25	47
7	85	26	45
8	83	27	43
9	81	28	41
10	79	29	39
11	77	30	37
12	75	31	35
13	73	32	33
14	71	33	31
15	68	34	29
16	66	35	26
17	64	36	24
18	62	37	22

SCORES FOR AGE GROUPS ON DIP TEST



SCORES FOR AGE GROUPS ON DIP TEST



TEST-RETEST RELIABILITY ±5 ITEMS (10%)
(Standard deviation of distribution
of difference in items correct
measured on test and
on retest)

END
4-18-68